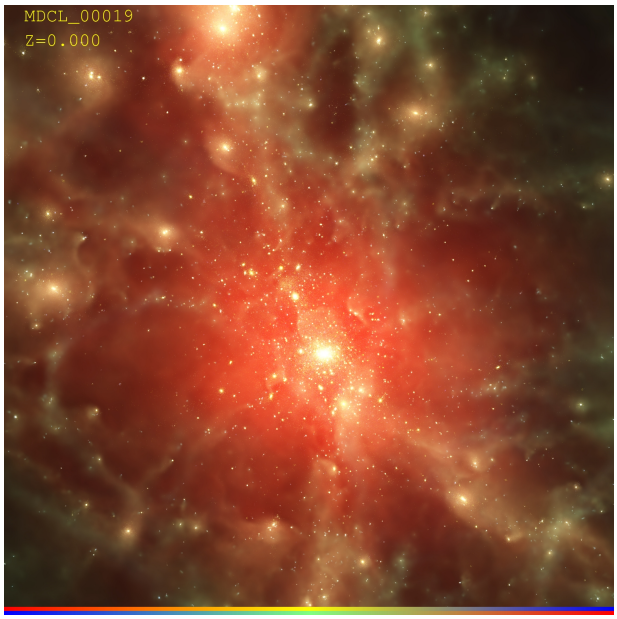
University of Nottingham
UK | CHINA | MALAYSIA

Cosmic Cartography: tracing the large-scale structure around galaxy clusters

Daniel James Cornwell



Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

If you believe in yourself, with a tiny pinch of magic, all your dreams can come true

– Spongebob Squarepants

Supervisors:	Prof. Alfonso Aragón-Salamanca	
	Prof. Meghan Gray	
	Prof. Frazer Pearce	
Examiners:	Prof. Scott Kay	(University of Manchester)
	Dr. Simon Bamford	(University of Nottingham)
Submitted:	24 May 2024	
Examined:	5 July 2024	
Final version:	28 September 2024	

Contents

Abstract

On the largest cosmological distance scales, the cosmic web forms the backbone of the Universe. This network connects dark matter, galaxies and gas through a variety of structures. Galaxies that reside in different cosmic web environments are subject to a range of different physical mechanisms which can ultimately dictate the evolutionary path they take. At one end of the density scale, galaxy clusters exist that can quench and ultimately transform galaxies as they encounter the intracluster medium and the high density of galaxies. This results in a higher fraction of quiescent and early type galaxies in clusters compared to the field. Further down the density scale are cosmic filaments. A substantial fraction of the Universe's mass budget is located in these environments. Filaments are a conduit through which galaxies are funneled through into the dense core of a galaxy cluster. Whilst there is strong evidence suggesting that galaxies experience environmental influence long before they reach the core of a cluster, (pre-processing), exactly where and when these changes take place is not well constrained.

Motivated by upcoming wide-field spectroscopic surveys of galaxy clusters, such as the WEAVE Wide Field Cluster Survey (WWFCS), we begin with an investigation into the feasibility of extracting cosmic web filaments around galaxy clusters with surveys like the WWFCS. We use hydrodynamic simulations from TheThreeHundred project of galaxy clusters to design mock observations for the WWFCS by taking into account observational selection effects. After extracting cosmic web filaments around galaxy clusters using the topological structures extractor `DisPerSE` in our mock observations, we compare them to the 'ground truth' simulated case. Reassuringly, we find that surveys like the WWFCS will succeed in providing detailed maps of the local cosmic web around galaxy clusters.

We then turn our attention to one of the main questions we aim to answer in this thesis: can we accurately allocate individual galaxies to different cosmic web environments? This is a crucial test: for surveys like the WWFCS to draw conclusions on how galaxy evolution is impacted by its environment, we need to know how accurately we can allocate galaxies to these environments in the first place. In order to answer this question, we design a framework where we separately use information available from simulations and from observations to classify galaxies as residing in the cluster core, filaments, and those that located in neither. Whilst we report galaxy-to-galaxy and cluster-to-cluster variations in our ability to classify galaxies, overall, characterizing filamentary galaxies is highly uncertain. However, most importantly, we outperform a random classification of galaxies to different environments. Taking into account our statistical treatment, surveys like the WWFCS will be able to draw robust conclusions on how environment shapes galaxy evolution. We also briefly discuss more recent work using Machine-Learning to drastically improve our success rates.

Having established our ability to accurately map cosmic filaments in the infall regions of massive galaxy clusters, we then consider the systematic identification of galaxy groups. Like filaments, evidence suggests that galaxy groups can preprocess galaxies prior to their accretion onto a cluster. Given that cosmic web nodes are directly identifiable through observations, we test their co-location with galaxy groups. In this study, we show a substantial fraction of massive groups that are distant from the cluster core match to cosmic web nodes. `DisPerSE` can therefore be used to identify both cosmic web filaments and galaxy groups with relative success.

The work in this thesis uses simulations from TheThreeHundred project to test, optimize and forecast the ability of upcoming surveys to trace the cosmic web around galaxy clusters. Along with previous studies, we pave the way for wide-field optical spectroscopic surveys of galaxy clusters to extract cosmic filaments and galaxy groups.

Acknowledgements

Throughout my PhD there have been many highs and lows – I want to acknowledge the people most responsible for these highs. I firstly would like to thank my official and unofficial supervisors for their guidance and support over these last four years. Alfonso; for your encouragement, mentorship and genuine enthusiasm about our work. I also thank you for giving me opportunities to ask you stupid questions and sharing the burden when things aren't going so well, as well as celebrating with me when things are looking up. Meghan; for your supportive personality, guidance in working between observations and simulations and your ability to think about the bigger picture. Frazer; for your constant stream of new ideas and thoughts. Ulli; you have been a constant source of encouragement and education. I started my PhD not knowing anything about the topic in this thesis and you have been a huge source of inspiration. I appreciate all you have done: arranging impromptu meetings with me so you can explain concepts, standing up for me in meetings when I cannot find the right words to say and, most importantly, educating me about Austrian cuisine. I am genuinely fortunate to have such supportive supervisors and couldn't have done it without you.

I'd like to thank my collaborators, for your input into my work as well as career advice. A special thanks goes to Alexander Knebe, Jose Alfonso Lopez Aguerri, Weiguang Cui, Tristan Weaver, Han Aung and Daisuke Nagai: I've really enjoyed working with and learning from you all. I am very grateful for Phil Parry – our resident I.T./tech guru. Phil has fixed more of my software/hardware problems than I can count and didn't even bat an eyelid when, on numerous occasions, I locked myself and others out of Captain by overloading it. I'd like to say a special thank you to Margaret for our morning chats in the kitchen and for making CAPT a better place.

Thanks to all my fellow PhD students. CB and Josh, for our frequent cafe visits when we were all panicking about our theses, (thesises, faeces?). Karel, Jen and Matt for coming to pizza pals (PPs) and not complaining about my attempt at home made pizza. Tomas, Steph, Kellie, Lizzie, Jimi, Harry, Mikey, Joe and Fiona and so many others – thank you for making my time at Nottingham enjoyable and worthwhile. Thank you to everyone that has come to our Eurovision, Birthday and Halloween parties and for carrying me during pub quizzes. I thank my friends Ryan, Tom, Cassie, Harrison and Katie for our ridiculous conversations

and providing me with distractions from work when I've most needed it.

Thank you also to my Mum, Dad, Sister, Sophia, my Grandparents and my Nan. Thank you for showing genuine interest in my work, supporting me in numerous ways over these last four years, (and for the 22 before it), and for providing me with the resources that have been responsible for me getting to this position. I am incredibly lucky to have you all, even if we're all a bit silly in our own way. Daisy – my cat – even though you genuinely enjoy attacking me, I LOVE YOU, no matter how *scronky* you are. Finally, I want to thank Kira. You mean the world to me. Through all the highs and all the lows, coming back home to you makes it all worthwhile.

Published Works

Almost all of the content in this thesis has already been published in the following three works:

1. **Cornwell D.**, Kuchner, U., Aragón-Salamanca A., Gray, M., Pearce, F., Aguerri, J., Cui, W., J. Méndez-Abreu, Peralta de Arriba, L., Trager. S, C., 2022. *Forecasting the success of the WEAVE Wide-Field Cluster Survey on the extraction of the cosmic web filaments around galaxy clusters.*. Monthly Notices of the Royal Astronomical Society, 517, 1678 .

2. **Cornwell, D.**, Aragón-Salamanca, A., Kuchner, U., Gray, M., Pearce, F., Knebe, A., 2023. *The probability of identifying the cosmic web environment of galaxies around clusters motivated by the Weave Wide Field Cluster Survey.* Monthly Notices of the Royal Astronomical Society, 524, 2148.

3. **Cornwell, D.**, Kuchner, U., Gray, M., , Aragón-Salamanca, A., Pearce, F., Cui, W., Knebe, A., 2023. *The localization of galaxy groups in close proximity to galaxy clusters using cosmic web nodes.* Monthly Notices of the Royal Astronomical Society, 527, 23

4. Weaver, T., Aung, H., **Cornwell, D.**, Nagai, D., Aragón-Salamanca, A, 2023. *Reconstructing Filaments around Galaxy Clusters from Spectroscopic Surveys using Machine Learning.* Research Notes of the American Astronomical Society, 7, 268

Chapter 3 is based on material published in Paper 1, Chapter 4 is based on Paper 2, Chapter 5 is based on Paper 3 and Section 6.2.2 is based on Paper 4.

The vast majority of the work presented in this thesis was carried out by the author, with supplementary advice from various co-authors listed above. In any instances where the work includes the product of larger collaborations, this is explicitly mentioned in the relevant Chapter.

Chapter 1

Introduction

The ultimate aim of this thesis is to try and better understand our ability to extract cosmological structures in close proximity to galaxy clusters. If we want to eventually understand the role that different cosmic web environments play on the evolution and formation of galaxies, then we need to investigate whether we can accurately classify galaxies into such environments. In doing this, we need to understand what exactly we mean by the cosmic web, as well as realizing the individual components that comprise it. Only after laying down this groundwork can we properly investigate their dependencies, and, in turn, the role that large-scale structure plays in the mass assembly of galaxies.

In this chapter, we first introduce the cosmic web and its detection in Section 1.1. In Section 1.2, we discuss the individual components of the cosmic web and methods that are used to detect them observationally. In Section 1.3, we talk about the link between different cosmic web environments and the evolution of a galaxy residing in them. Finally, in Section 1.4, we describe the structure of this thesis.

1.1 Large-scale structure: the cosmic web

The Universe we live in exhibits complex structures across a wide range of scales. At some of the smallest scales, quantum mechanics dictates the presence and structure of subatomic particles that form nuclei. The atoms these comprise form the

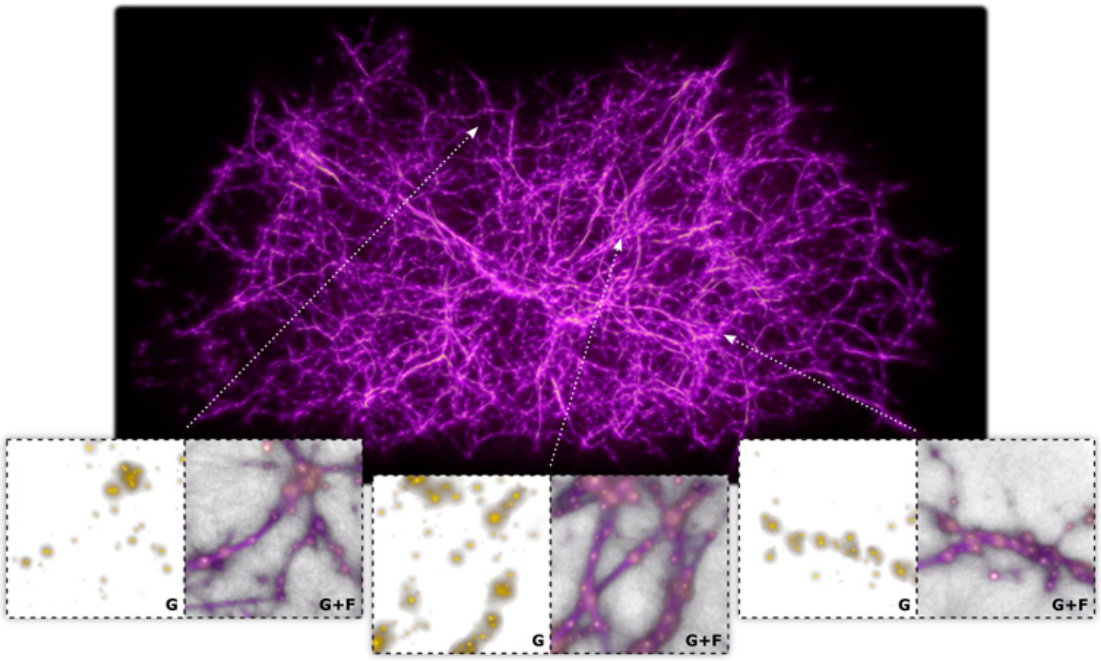


Figure 1.1: Cosmic web reconstruction using galaxies from the Sloan Digital Sky Survey (SDSS). Top panel: the filament network created using the Monte Carlo Physarum Machine, taking inspiration from the slime mold model. The bottom panels shows three regions with the underlying SDSS galaxy positions and the superimposed filament density field. The model can trace the underlying galaxy distribution, taking in only galaxy coordinates, redshifts and masses. Taken from Burchett et al. (2020).

building blocks of chemicals, compounds and organisms. The neuronal network is an example of such a structure that exists on a human scale. This is a hierarchical network in which neurons cluster into circuits, columns and different interconnected functional areas (Vazza & Feletti, 2020). Nature has a way of replicating these networks on larger scales. The 'slime mold' *Physarum polycephalum* is a unicellular organism that is known to explore its environment for food sources and shape itself into highly intricate networks. The slime mold has been used to reconstruct highly efficient transport networks, such as the Tokyo rail system (Tero et al., 2010). However, the multi-scale nature of these networks extends further. Within galaxies, filamentary structures extend on parsec[1] scales that connect giant molecular clouds (Jackson et al., 2010). On the largest scales, the Universe weaves a vast network known as the cosmic web that connects different cosmic structures.

[1]The parsec, abbreviated as pc, is a unit often used for measuring cosmological distances. $1\text{pc} = 3.09 \times 10^{16}\text{m}$

Whilst the cosmic web spans many more orders of magnitude in distance compared to the neuronal networks of the brain and slime mold, it demonstrates a remarkable similarity to these smaller-scale networks (e.g. Burchett et al. 2020; Vazza & Feletti 2020). This implies a scale-free nature to these networks. Figure 1.1 illustrates how the 'slime mold model' is used to reconstruct the cosmic web. In what follows, we describe the largest of these types of network: the cosmic web.

1.1.1 History of the cosmic web

Whilst Zel'dovich (1970) was the first to predict the presence of components of the cosmic web, (namely pancakes and voids), the term 'cosmic web' was first coined by Bond et al. (1996). They found that the final-state cosmic web is actually present in embryonic form in the overdensity pattern of the initial fluctuations. This structure arises due to the presence of small perturbations that propagate through the early Universe's primordial plasma, which is visible through temperate fluctuations in the cosmic microwave background. This results in over and underdensities, providing the seeds of structure growth (Springel et al., 2006). Over cosmic time, these fluctuations are amplified through gravity and build highly asymmetrical structures. Overdense regions firstly collapse to form walls, then collapse through two principal axes to form filaments before finally forming clusters (Arnold et al., 1982). The end result is the cosmic web – the dark matter skeleton that connects the Universe.

1.1.2 Observing the cosmic web

At a similar time to the theoretical predictions of the cosmic web (Zel'dovich, 1970; Shandarin & Zeldovich, 1989), some of the first large-area galaxy redshift surveys provided observational evidence of a local anisotropic distribution of matter. The first CfA redshift survey, conducted by Huchra et al. (1983), mapped 2400 galaxies down to $m \approx 14.5$ using optical spectroscopy. In order to map galaxies in three-dimensions, we need an estimate of their line-of-sight distance. From spectroscopy one can measure redshift, via the relative change in wavelength of a spectral feature. In an expanding Universe, the redshift is related to the distance

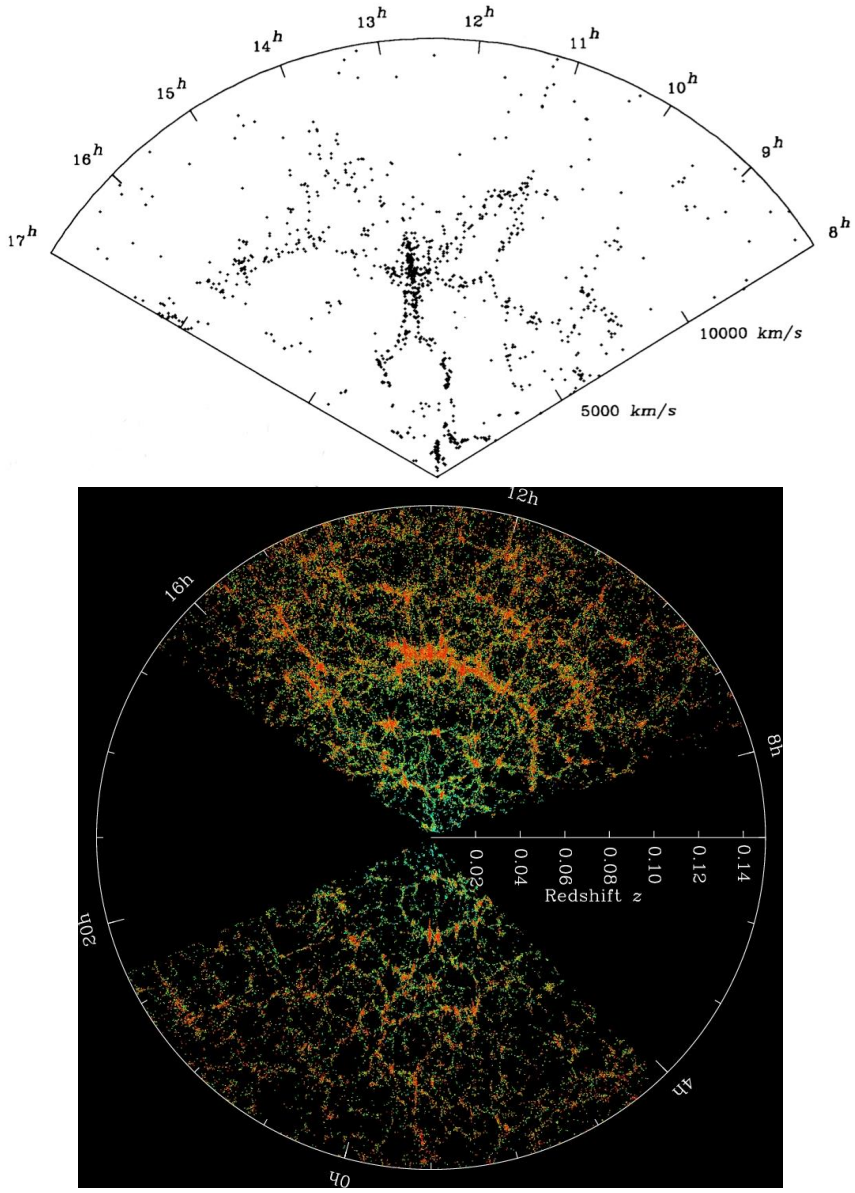


Figure 1.2: Comparison of the first redshift survey to provide evidence of a large-scale network compared to a more recent, improved map. Top panel: approximately 2000 galaxies identified through spectroscopy in the first CfA survey (de Lapparent et al., 1989). Bottom panel: approximately 700,000 galaxy identified in the SDSS-MAIN survey (York et al., 2000), providing a much more detailed picture of the cosmic web.

to the source. For small cosmological distances, one can approximate distances using the recessional velocity inferred from the redshift, via the Hubble Constant. At greater distances, one needs to use a full cosmological model to convert redshift to a distance. However, there is a complication: in addition to the cosmological expansion, the observed redshift also includes components due to the peculiar velocities of both the observer and the emitter in their local cosmological rest frames. The correction for the observer's velocity affects all measurements in a systematic way. On the other hand, if we simply convert observed galaxy redshifts to distances, they will scatter around the true distance. For nearby galaxies, the observed redshifts are dominated by peculiar velocities. Further afield, distances estimated from redshift are relatively accurate, and the cosmic web can be discerned. However, the blurring along the line-of-sight complicates the identification of individual structures.

Subsequently, de Lapparent et al. (1989) complemented this survey and was the first to detect a network of structures, including clusters and voids. Over the following decades, numerous redshift surveys studied the cosmic cartography of the low redshift Universe. Surveys such as the Sloan Digital Sky Survey (SDSS, York et al. 2000), the Two-Degree Field Galaxy Redshift Survey (2dFGRS, Colless et al. 2001), the Galaxy and Mass Assembly survey (GAMA, Liske et al. 2015), the VIMOS Public Extragalactic Redshift Survey (VIPERS, Mohammad et al. 2018) and, more recently, the Dark Energy Spectroscopic Instrument (DESI, Secco et al. 2022) all probe the distribution of galaxies over a wide area, providing compelling observational evidence for the presence of a cosmic web. Over the last fifty years, the detail of the map of the local Universe has improved tremendously, made possible by higher fiber multiplexes resulting in improvements in sampling rates of galaxies as well as the exploration of fainter magnitudes. This is illustrated in Figure 1.2, where a much clearer depiction of the large-scale structure of the Universe is presented by the SDSS in the bottom panel, compared to the CfA's 'slice of the Universe' distribution in the top panel.

Through the aforementioned surveys, we are now able to visualize some of the striking features present in the cosmic web. These features include nodes, (often identified as clusters), long tendrils connecting nodes, known as cosmic filaments, sheets (or walls), and void regions. We will go into more detail about these structures in the following sections. The presence and evolution of these four

structures follows a well-defined process, described by the Zel'dovich formalism. For a three-dimensional overdense region to collapse, its self-gravity must overcome both the cosmological expansion and the tidal field due to surrounding structure. This requirement may be reached in some directions, but not in others, so a given region may only be collapsing along one or two axes. Generally, a three dimensional overdense region collapses along all three directions under the force of gravity. This initial density is not static, but is undergoing cosmological expansion. However, the collapse is primarily along one direction, resulting in the formation of a two dimensional sheet. Following this, the collapse continues through the next most significant principal axes to form filaments before finally forming clusters (Arnold et al., 1982). This well ordered sequence of anisotropic gravitational collapse is illustrated in Figure 1.3. We note that this is a simplified model that we are imposing on a more complicated reality to aid understanding.

Currently, observations alone can not provide a holistic view of the formation and evolution of the cosmic web. One of the main reasons is that even at low redshift, certain low-contrast features of the cosmic web remain incredibly difficult to distinguish from spurious patterns in the matter distribution. Cautun et al. (2014) displayed the distributions of overdensities in different cosmic web structures and found that there are large overlaps in the density distributions between each consecutive environment. For example, filaments have a broad overdensity distribution which overlaps with nodes, walls and voids. As a result, characterizing the cosmic web is non-trivial, especially observationally. Fortunately, there are other tools we can use that allow us to help plan and optimize the characterization of cosmic structures in future observations.

1.1.3 Simulations of the cosmic web

It is important to note that our current understanding of the large-scale structure of the Universe requires the presence of dark matter. Whilst we do not know what dark matter is made of, it is theorized to be the underlying skeleton of the cosmic web, which baryonic matter is later accreted on to. Cold dark matter (CDM) is the most widely accepted paradigm. Simulations have shown that, without the incorporation of CDM, baryonic matter would not have sufficient time to form the structures we see in the present day (Davis et al., 1985). Furthermore, the most

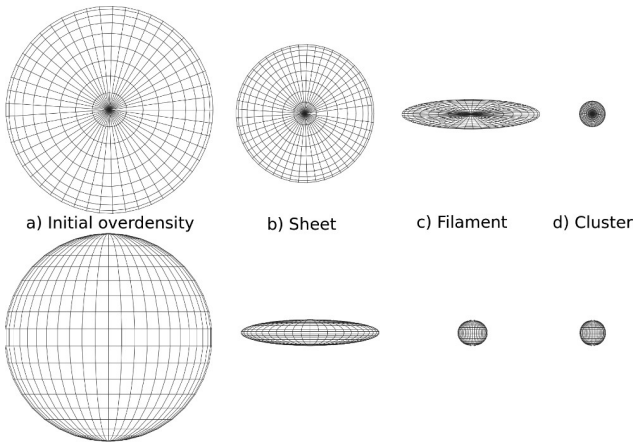


Figure 1.3: The formation and evolution of cosmic web structure, from an initial overdensity to a galaxy cluster. The sequence starts in the left most panel where an ellipsoidal overdensity collapses along one axis to form a sheet. This is followed by a contraction along the second major axis to form a filament before finally fully collapsing to form a cluster. This figure is taken from Cautun et al. (2014).

commonly used cosmological model also incorporates 'dark energy' (Λ) to explain the accelerating expansion of the Universe (Riess et al., 1998; Perlmutter et al., 1999). These contents form the basis of the 'standard model' of cosmology, known as ΛCDM.

In terms of numerical simulations of large-scale structure, not knowing the nature of dark matter is, counter-intuitively, currently not a huge hindrance. CDM is collisionless and only interacts through gravity, meaning that computer simulations ran only using dark matter particles are relatively inexpensive. Given that evidence suggests that CDM outweighs baryonic matter by a factor of five to one, dark matter only simulations are relatively successful at recreating observed large-scale structure, (see Angulo & Hahn 2022 for a review), as well as the underlying substructure of dark matter haloes (Kuzio de Naray et al., 2009). Also known as N-body simulations, these involve using the equations of motion of a large number of particles, as well as their interacting forces, and numerically integrating them over time to study the evolution of a system from a set of initial conditions.

The Universe, however, does not only contain dark matter, and a lot of the astrophysical processes that play an important role in the evolution of cosmic structures are directly impacted by baryonic physics. As a result, to make N-body simula-

tions more realistic, it is necessary to model with hydrodynamics. By simulating with hydrodynamics, we can incorporate mechanisms that include baryonic, gas and stellar physics. These processes are mainly confined to smaller, kpc scales, such as active galactic nuclei (AGN) feedback, supernova feedback, radiative cooling and shocks (Springel et al., 2001; Springel, 2005; Beck et al., 2015; Hopkins, 2015; Davé et al., 2019). Whilst including these features in the simulations is a necessary component to understand galaxy formation and evolution, they are far more computationally expensive than the dark matter only N-body simulations.

There currently exists a plethora of simulations that focus on different components of the cosmic web. These range from large-scale cosmological box simulations to zoom-in resimulations of overdense regions, such as galaxy clusters. I discuss the simulations used in this work in detail in Chapter 2.1.2.

1.2 Components of the cosmic web

1.2.1 Galaxy clusters

Galaxy clusters are the densest region in the cosmic web that typically reside at the location of nodes (Aragón-Calvo et al., 2010; Cohn, 2022). Being the most massive gravitationally bound structures in the Universe, galaxy clusters host hundreds to thousands of galaxies and are generally differentiated from other structures by hosting a halo with mass exceeding $10^{14} M_{\odot}$. They are virialized objects, meaning that the kinetic energy of the cluster galaxies and intracluster gas is balanced by the gravitational potential energy of the entire system. Galaxy clusters are theorized to form via hierarchical structure formation. This is a 'bottom-up' process, in which small objects merge early on to form the most massive structures that we identify today (Gunn & Gott, 1972).

Dark matter haloes do not have sharp boundaries. Instead, it is common practise to use the following equation to define a radius based on the overdensity:

$$r_{\Delta} = \left(\frac{3 M_{\Delta}}{4 \pi \Delta \rho_{crit}} \right)^{1/3}, \tag{1.1}$$

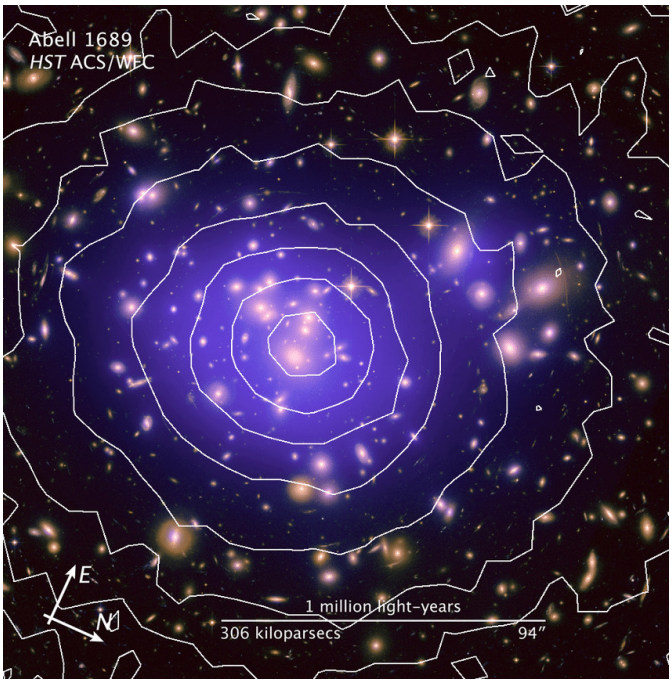


Figure 1.4: Observations of the galaxy cluster Abell 1689. Overlaid on a composite Hubble Space Telescope image is the dark matter distribution, inferred through lensing, in blue. The white contours correspond to the spatial distribution of mass from Chandra X-ray data that traces the ICM in the galaxy cluster. Image credits: NASA, ESA, E. Julio., (Scarlata et al., 2013).

where Δ is the density contrast ($(\rho - \bar{\rho})/\bar{\rho}$), M_Δ is the mass enclosed within a given radius r_Δ and ρ_{crit} is the critical density of the Universe. The values of Δ differ depending on the types of observations one is using and what part of the cluster region they are probing. Going forward, we will refer to R_{200}, that is the radius r at which the density drops below $200\rho_{crit}$. Simulations show that this radius is sufficient to separate infalling material from the relaxed part of a cluster. Typically, for galaxy clusters, this is between 1-3 h^{-1} Mpc. h is the dimensionless equivalent of the Hubble Constant H_0. Typically $h = H_0$ / (100 km/s/Mpc), but other definitions are found. Galaxy clusters are complex bodies and feature many different components, most of which are directly or indirectly observable. In the next few paragraphs, we briefly outlay the constituents of a galaxy cluster.

Firstly, is the dark matter halo: the galaxy cluster's largest mass component, contributing to approximately 90% of the total cluster mass (Gonzalez et al., 2013). At low redshift, galaxy clusters have decoupled from the Hubble expansion and contain gravitationally bound matter, which, in turn, is responsible for the

accretion of baryonic matter. Whilst it is not possible to observe the dark matter halo, indirect methods, such as gravitational lensing yield information about its mass content and density (Soucail et al., 1988; Hoekstra et al., 2013). This is illustrated in Figure 1.4 where the dark matter content of Abell 1689 is illustrated in blue. Further to this, N-body simulations indicate that the radial density profile of a dark matter halo can be modelled well by the Navarro-Frenk-White profile (NFW; Navarro et al. 1997). Although problems exist with this model, it is generally a good approximation and is further supported by weak lensing measurements (Okabe et al., 2013).

The cores[2] of galaxy clusters are high density regions, where a large number of galaxies are packed into a relatively small volume of space. Within the virialized core of the cluster is the Intracluster Medium (ICM), contributing to approximately 10% of the cluster's mass budget (Lau et al., 2009). The ICM is the main baryonic component of a galaxy cluster. It is a relatively low density (10^{-3} cm^{-3}) hot (10^{7} K) plasma that permeates the cluster main halo and is visible at X-ray wavelengths (Evrard, 1990; Mohr et al., 1999; Peterson & Fabian, 2006) and through the Sunyaev-Zeldovich effect (Sunyaev & Zeldovich, 1972). The contours in Figure 1.4 illustrates the X-ray emission from the ICM. Whilst the origin of ICM is debated, a common theory is that ICM is heated through various hydrodynamic processes, such as AGN and supernova feedback. In turn, this is enabled through continuous accretion of matter from cosmic filaments and through merger events (Di Mascolo et al., 2023).

The brightest galaxy in a cluster is known as the BCG (brightest cluster galaxy) which theoretically sits at the bottom of the cluster's potential well and is surrounded by lower mass satellite galaxies. The BCG is, therefore, often used as a reference for the 'centre' of a galaxy cluster, which aligns well with the peak of the X-ray emission (Lin & Mohr, 2004), although evidence suggests that this can depend on the recent accretion history of the cluster (Martel et al., 2014). BCGs are thought to form through the merger of several massive galaxies early in the cluster's history, as well as the ingestion of smaller satellite galaxies (Ostriker & Hausman, 1977; Merritt, 1984). They are generally elliptical and are some of the most massive galaxies in the Universe, sometimes exceeding stellar masses of $M_* > 10^{12} M_\odot$ (Zhao et al., 2015). Within the inner few hundreds of kiloparsecs of

[2]Throughout this thesis, we refer to the 'core' as the virialized region of a cluster

the cluster exists Intracluster Light (ICL). ICL is hypothesized to be the light of stars that have been stripped from interactions between cluster galaxies and the ICM and traces the potential of the dark matter halo (Lin & Mohr, 2004). Whilst the combination of the luminosity originating from the BCG and ICL contain a significant fraction of the cluster's total luminosity, together, they contribute to less than 1% of the overall cluster's mass (Gonzalez et al., 2013).

1.2.1.1 Methods for observing clusters

Now that we have discussed the typical ingredients of a galaxy cluster, we layout common methods for observing them.

In 1958, George Abell published one of the first galaxy cluster catalogues, accumulating a total of 2712 clusters (Abell, 1957). After identifying a set of crtieria, he identified clusters visually by counting overdensities of galaxies that appeared in a stack of photographs from the Palomar Observatory. Since, many of these objects have been observationally verified as galaxy clusters and his catalogue is still used to this day. There are several methods that are commonly used for classifying and studying galaxy clusters. These include X-ray emission from the ICM (e.g. Sarazin 1986), galaxy-galaxy cluster gravitational lensing (e.g. Broadhurst et al. 1995), photometric analysis of the properties of clustered galaxies (e.g. Gladders & Yee 2000) and the spectral distortion of CMB photons by the hot ICM, known as the Sunyaev-Zel'dovich effect (Sunyaev & Zeldovich, 1972). In this thesis, we only focus on spectroscopy.

Optical spectroscopy and photometry are regularly used to infer the light profiles of galaxies in and around clusters. The flux from galaxies can be measured quickly over large wavelength ranges, known as color bands, (photometry), or over much smaller wavelength increments, (spectroscopy). Whilst there is much information that can be extracted using photometry, high-quality photometric redshifts alone are not sufficient for accurately determining cluster membership. Instead, the spectrum of a galaxy can provide a more robust determination of its redshift and its baryonic properties. Whilst this method has longer integration times and is more costly, at low redshift, the uncertainities in the redshift measurements are typically reduced from ± 0.1 to ± 0.001 (Hopkins, 2015). The error in the line-of-

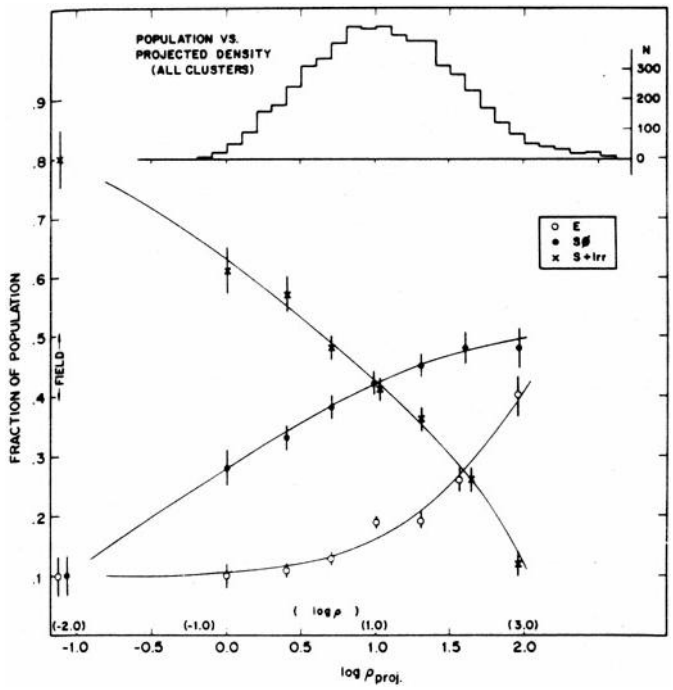


Figure 1.5: An illustration of the morphology density relation. This figure depicts some of the first observational evidence of the connection between the morphology of a galaxy and the density of its environment. It displays the fraction of different galaxy types, (elliptical, lenticular and spiral/irregular) in different projected densities. Taken from Dressler (1980).

sight distance, derived from Hubble's law, can be reduced from hundreds of Mpcs to tens of Mpcs, allowing a much more accurate cluster membership quantification. As a result, spectroscopic surveys of galaxy clusters and their members are preferentially used for detailed studies of galaxy cluster environments. Examples of such are the WINGS (Fasano, G. et al., 2006) survey and the GOGREEN survey (Balogh et al., 2017). Both of these surveys provide follow up spectroscopy of the cluster members from clusters previously identified in X-ray/by the SZ effect.

1.2.1.2 The environment of a galaxy cluster

Galaxy clusters are regions of significant astrophysical interest due to their extreme *environments*. They are unique laboratories for investigating how hydrodynamic and gravitational interactions may influence the evolution of a galaxy. Such processes include ram pressure stripping (Gunn & Gott, 1972), galaxy harassment

(Moore et al., 1998) and starvation (Larson et al., 1980), which can all act to influence the morphology and star formation rates of these galaxies. One of the first quantitative studies in this area was that of Dressler (1980), who discovered the so called 'morphology density relation'. As illustrated in Figure 1.5, in high density environments, early-type elliptical and lenticular galaxies are more commonly located. Conversely, in low density environments, galaxies are more likely to be late-type spirals. This finding implies a connection between galaxy evolution and their surrounding environment, providing early evidence for this relationship.

Beyond the core of the cluster is the cluster *infall region*. In hierarchical clustering, galaxy cluster sized haloes accrete smaller haloes over time. This results in the accreted objects having radially infalling velocities as well as velocities with large random motions (Diaferio & Geller, 1997). It is in this region where the gravitational influence of the cosmic web skeleton overcomes that of the cluster and hence, is a region of large complexity. Further to this, there is evidence that galaxies experience environmental effects in the infall region of clusters, a term coined *pre-processing* (Zabludoff & Mulchaey, 1998). This further has the potential to influence their intrinsic properties before they are accreted on to the cluster (Cortese et al., 2006). In the next few sections, we describe in detail the possible cosmological structures that provide the channels for such pre-processing.

1.2.2 Galaxy groups

Second to galaxy clusters, galaxy groups exhibit the largest dark matter haloes of all cosmic structures. The definition of a galaxy group varies throughout the literature, (see Lovisari et al. 2021 for a review on galaxy groups). Examples of groups include *Compact groups*: a small number of galaxies, usually between 3 and 10 (Taverna et al., 2023), that are in close proximity and are relatively isolated (Hickson et al., 1992). Hickson (1982) created a catalogue of compact groups in 1982, one of which is Stephan's Quintet (Stephan, 1881). *Fossil groups*: galaxies that are embedded in a giant X-ray halo, typical of a group, and are theorized to be the latest stage in galaxy evolution, as several group galaxies merge into one galaxy (Ponman & Bertram, 1993).

More massive galaxy groups with higher number densities are more likely to retain

an 'intragroup medium' and will be more influential in reshaping the properties of galaxies. Criteria for defining galaxy groups vary; some N-body simulations require a halo mass of $M_{500} > 10^{13} M_{\odot}$ (Le Brun et al., 2014), whilst other studies using spectroscopic observations suggest a halo mass of $10^{12.5} M_{\odot}$. The review on galaxy groups by Lovisari et al. (2021) suggested that astronomers use the 'rule of thumb' that systems with 50 galaxies or more are clusters, and less than this are groups. Either way, there is clearly a contention regarding the 'true' definition of galaxy groups as it varies depending on the scale of the study and the scientific motivation.

Throughout this thesis, we are focused on galaxy groups that are in close proximity to galaxy clusters and are strong candidates for presenting signs of pre-processing. Galaxy clusters accrete a significant fraction of their $z = 0$ galaxy population through galaxy groups (McGee et al., 2009) and, therefore, are interwoven into the large-scale cosmic web. Their presence in the infall region of clusters can significantly influence the galaxy-galaxy merger rates (Vijayaraghavan & Ricker, 2013) as they can easily sweep up field galaxies and grow quickly. Group members are also likely influenced by their group environment prior to their accretion onto the main halo of the cluster, shown in simulations (Bahé & McCarthy, 2014; Jung et al., 2018), and observations (Jaffé et al., 2016; Bianconi et al., 2017; Haines et al., 2018; Benavides et al., 2020; Lopes et al., 2023).

Currently, there exists a range of techniques that successfully detect groups in larger-scale observations. One of the most common is the Friends-of-Friends (FoF) percolation algorithm (Huchra & Geller, 1982), which has been used, for example, in the 2dFGRS (Eke et al., 2004), the SDSS (Berlind et al., 2006) and GAMA (Robotham et al., 2011). However, in the infall region of galaxy clusters, group-finding is a non-trivial task. Inside the cluster's potential well, galaxies have large random motions relative to one another. In addition, infall motions towards the cluster and filaments also dramatically perturb the galaxy distribution (Kuchner et al., 2021). Therefore, with respect to the observer, galaxies in the vicinity of clusters and groups may have similar distances, but their large random motions lead to very different redshifts. This manifests as long, artificially extended structures, known as the 'Fingers of God' (FoG; Tully & Fisher, 1978). The length of the FoG for a massive cluster with velocity dispersion of 1400 km s^{-1} corresponds to 20 h^{-1} Mpc extending in each direction (Kuchner et al., 2021). Therefore, in

the vicinity of galaxy clusters, we are limited to more laborious, non-systematic methods of group detection that are used on a cluster-by-cluster basis. For example, previous studies have relied on visually inspecting 3D maps of the galaxies in RA, DEC, and redshift space to detect possible galaxy overdensities (Jaffé et al., 2013) or the Dressler-Shectman test that compares the local velocity and velocity dispersion for each galaxy against a global value (Dressler & Shectman, 1988). These and other methodologies are successful on a single cluster basis but become very time consuming when considering multiple clusters, where hundreds of groups are observed. We address this problem in Section 5.

1.2.3 Cosmic filaments

Cosmic filaments are arguably the most striking feature in the cosmic web. Since $z \approx 2$, cosmic filaments host half of the mass budget of the Universe, despite only contributing to 6% of its volume (Cautun et al., 2014). Their ubiquity is evident in Figure 1.6, where the cosmological box is connected through long ridges known as filaments. This same figure shows the remarkable multi-scale nature of these features. On the largest scales, filaments can be seen connecting the bright peaks in the density field, where galaxy clusters are located at their intersection (Aragón-Calvo et al., 2010). The bridges that connect clusters to clusters are often tens of megaparsecs in length, as found in observations (e.g. Finoguenov et al. 2003) and in simulations (e.g. Galárraga-Espinosa et al. 2022). In the middle and lower panel of Figure 1.6, a 100 and 20 h^{-1} Mpc cut out of the native simulations are displayed. It is here where we can see an increase in the number of thinner cosmic filaments that feed the core of a galaxy cluster, as seen as the bright peak in the center of the zoom-in plots. These filaments are typically several megaparsecs in length, as found in observations (e.g. Tanimura et al. 2020) and simulations (e.g. Rost et al. 2021) and act as highways that funnel galaxies and groups into the virialized cluster core.

Whilst there have been several studies on the multiscale length of filaments, their thickness is much more difficult to quantify. As previously mentioned, filaments are a later evolutionary state of cosmic walls and are therefore, a constantly evolving component of the cosmic web. Subsequently, filaments do not have definitive boundaries, and their thickness depends on the type of study and definitions used.

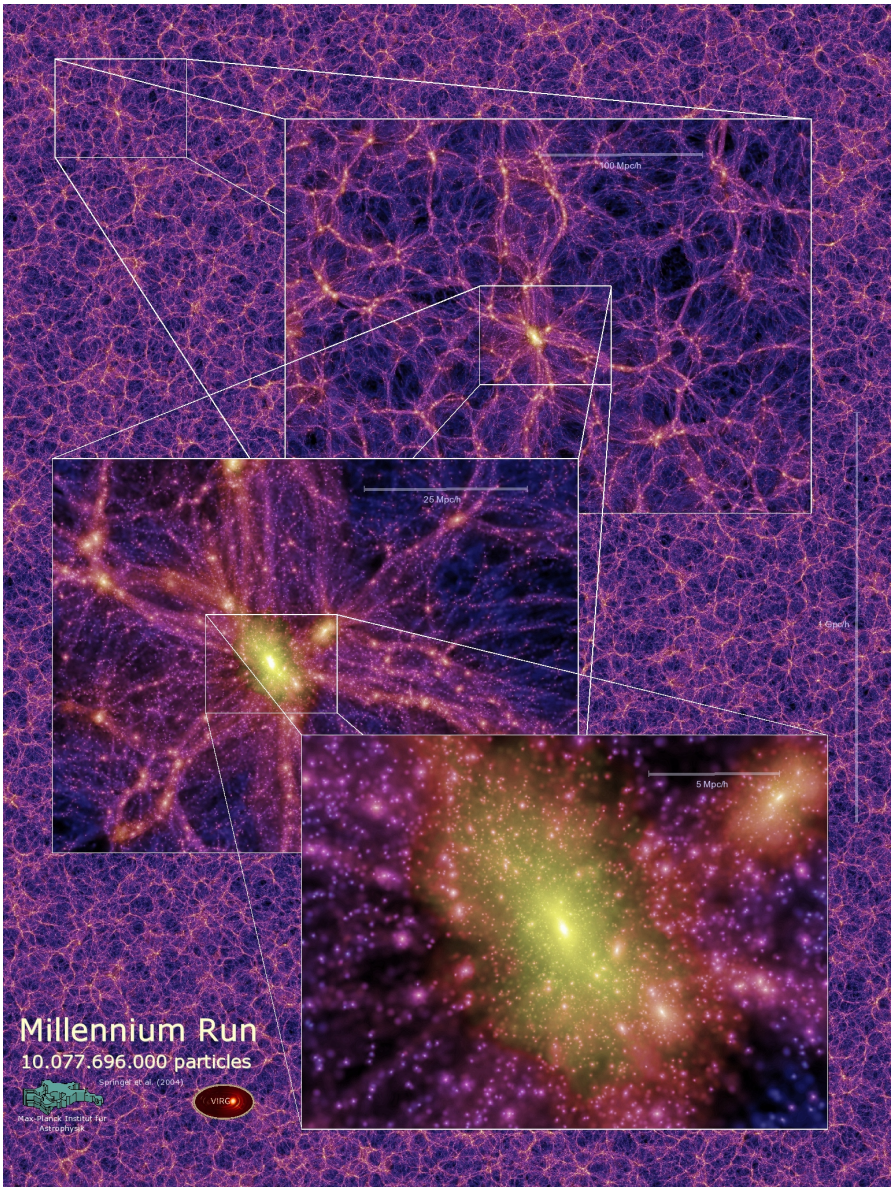


Figure 1.6: An illustration of the multi-scale nature of cosmic filaments in the larger-scale cosmic web. This picture is taken from the Millennium simulation (Springel et al., 2005), a cosmological hydrodynamic simulation. The first box shows the full simulation box, then each consecutive panel zooms in on the same region, showing a variety of distance scales, from the cosmic web to high density galaxy clusters that are connected by cosmic filaments.

Studies have shown that their radii can typically lie anywhere between 0.7 and 5 h^{-1} Mpc (Kuchner et al., 2020; Dolag et al., 2006). Further to this, there is evidence that suggests that filaments do not have a uniform thickness, such that they generally get thicker closer to nodes of the cosmic web (Rost et al., 2021).

The complex environment of cosmic filaments is further evidenced through the ambiguity in their detection. Observationally, on the largest scales, there are currently two main methodologies used to characterize cosmic filaments. The first is from optical spectroscopy of galaxies through redshift surveys, which is possible due to galaxies tracing the dark matter skeleton of the cosmic web. Simply put, with spectroscopy, precise redshifts, (and therefore, distances), can be calculated, allowing a robust estimation of the 3D position of galaxies, with the largest uncertainty in their line of sight distance. In tracing the positions of a large number of galaxies, the density field can be calculated, from which there are a variety of methods to detect the spine of a cosmic filament, as outlined in a review by Libeskind et al. (2017), and we further discuss in Chapter 2.3. Examples of catalogues of filaments derived via the aforementioned survey type are: Tempel et al. (2015); Alpaslan et al. (2016); Malavasi et al. (2016).

The second main methodology developed to tackle the mapping of cosmic filaments is through stacked X-ray emission of the warm hot intergalactic medium (WHIM). Numerical simulations indicate that a substantial fraction of baryons are found in filaments in the form of a hot plasma, known as the 'Warm Hot Intergalactic Medium' (WHIM; Aragón-Calvo et al. 2010; Cautun et al. 2014; Martizzi et al. 2019). However, due to the diffuse nature of filaments and their low contrast in density, it is often required to stack filaments in order to derive a statistically significant signal. Identification of cosmic filaments through X-rays also firstly relies on the detection of filaments through alternative means, namely optical spectroscopy. Furthermore, it is required to mask out the hot plasma emission from galaxy groups and clusters, adding further complexity to their detection. Finally, using the predefined catalogue of filaments, one would record the X-ray count around a certain distance to the spine of the filament in the appropriate energy band(s). Examples of filamentary detections through stacked X-ray emission are: Tanimura et al. (2020); Vernstrom et al. (2021).

The majority of work carried out in this thesis focuses on the detection of cos-

mic filaments in the infall region of galaxy clusters, in preparation for upcoming wide-field spectroscopic surveys. In this regime, there are a range of additional complexities that add significant difficulty to detecting cosmic filaments, even with the capabilities of next generation wide-field multi-object spectroscopic surveys. We further test and optimize the observational detection of cosmic filaments in Chapter 3 & 4.

1.3 Pre-processing

There is an ever increasing amount of evidence suggesting that the cluster environment is paramount in affecting the evolution of galaxies, as we alluded to in Section 1.2.1, and is most notably manifested in the morphology density relation. However, the role that separate cosmic web environments play in the evolution and mass assembly of galaxies in the infall region of clusters is much less well understood. Recent studies show that the transition of galaxies from 'field-like' to 'cluster-like' can occur at 2-3 virial radii (Haines et al., 2015, 2018; Kuchner et al., 2017; Bianconi et al., 2017; Tawfeek et al., 2022; Werner et al., 2021). This implies that galaxies in groups and filaments in the infall regions of clusters experience gravitational and/or hydrodynamical interactions during their infall, far before they traverse the cluster environment.

With respect to galaxy groups, research suggests that galaxies that sit in groups in the infall region of clusters have a reduced star formation compared to galaxies outside of groups in the same region (Bianconi et al., 2018). Vijayaraghavan & Ricker (2013) found that galaxies in groups experience pre-processing in the form of the removal of their hot halo and increased galaxy-galaxy merger rates. Cortese et al. (2006) also demonstrated that, in this region, galaxies in groups can suffer ram pressure stripping as well as starburst periods from galaxy-galaxy mergers.

With respect to cosmic filaments, observational and simulation based studies suggest that galaxies that reside closer to the spine of a filament are typically more massive, redder and more star forming (Alpaslan et al., 2016; Malavasi et al., 2016; Chen et al., 2017; Kraljic et al., 2017; Ganeshaiah Veena et al., 2018; Laigle et al., 2017; Sarron et al., 2019). Whilst this suggests that the filamentary structures

typically act to inhibit the supply of gas to galaxies, there is also evidence for a HI enhancement near filaments of the cosmic web (Kleiner et al., 2016; Vulcani et al., 2019), implying that galaxies are able to accrete cold gas from the filamentary environment. Both of these results point towards the presence of pre-processing and is therefore, a main target of upcoming wide-field observational surveys.

The infall region of a galaxy cluster is evidently a complex region. Coincidentally, much of our knowledge of the physical mechanisms at play here are limited by a lack of detailed spectroscopic observational data. In this thesis, we describe some of the final steps that are taken in preparing for observational surveys that are aimed at tackling these kinds of problems. After all, if we want to understand how cosmic web environments influence galaxy evolution, can we actually robustly classify them?

1.4 Structure of this Thesis

This thesis is structured as follows. In Chapter 2, we describe the simulation dataset that is used in this work. We also describe the astronomical survey motivating this work and the technical details of software used in this thesis. Chapter 3 presents our investigation into the feasibility of extracting cosmic web filaments after applying survey-like selection effects to mock galaxy clusters. After applying these selection effects, we investigate the reconstruction of the cosmic web around galaxy clusters by calculating different properties of the filamentary networks. In Chapter 4 we build on the research presented in Chapter 3 and investigate the allocation of mock galaxies to different cosmic web environments, using hydrodynamic simulations. We demonstrate the difficulty in allocating galaxies to filamentary environments and the need for large statistical samples of galaxies in filaments. In Chapter 5, we present a theoretical study into the detection of galaxy groups in close proximity to galaxy clusters using a topological structures extractor. Finally, the conclusions of this thesis, a summary of ongoing/future work and applications of this thesis to future observations are discussed in Chapter 6.

Chapter 2

Simulation Data and Technical Analysis

This Chapter describes in detail the simulation data and software employed in this thesis. Firstly, and most importantly for this work, we discuss the technical details of the simulation datasets in Section 2.1. In Section 2.2 we describe the survey that motivates the studies carried out in this thesis. In Section 2.3, we discuss the cosmic web extractor used in this work.

2.1 Simulations

We begin this section by describing the simulation dataset that forms the foundation of the work carried out in this thesis. In preparation for observations, we can use state of the art dark matter and hydrodynamic simulations to better constrain cosmic web environments. All of the work carried out in this thesis uses the TheThreeHundred galaxy cluster simulations (Cui et al., 2018, 2022). These are a set of resimulations of cluster regions from a larger dark matter only box: the MDPL2 Multidark Simulation (Klypin et al., 2016). We describe these simulations and the halo selection algorithim in Sections 2.1.1 - 2.1.3.

2.1.1 MDPL2 MultiDark simulations

The MultiDark simulations[1] are a widely used set of simulations. These simulations range in box size and resolution, from a comoving length of $4h^{-1}$ Gpc to $0.25h^{-1}$ Gpc (Klypin et al., 2016). The work carried out in this thesis is built on the MDPL2 MultiDark Planck simulation. MDPL2 is a periodic cube of comoving length $1\,h^{-1}\,\mathrm{Gpc}$ containing 3840^3 dark matter particles, each with mass $1.5 \times 10^9\,h^{-1}M_\odot$. MDPL2 uses *Planck* cosmology ($\Omega_\mathrm{M} = 0.307, \Omega_\mathrm{B} = 0.048, \Omega_\Lambda = 0.693, h = 0.678, \sigma_8 = 0.823, n_s = 0.96$). The Plummer equivalent gravitational softening length is $7h^{-1}$ kpc.

As MDPL2 is a dark matter only simulation, it is possible to simulate large cosmological volumes as the particles are only subject to gravitational interactions with each other. The initial conditions are seeded at a redshift of 100 using the Zeldovich approximation. The MDPL2 simulation then uses the L-Gadget-2 cosmological code (Springel, 2005), which is a modification of the widely used Gadget-2, a smoothed particle hydrodynamics code (SPH). Halo catalogues were then extracted using a range of halo finders, one of which we describe briefly in Section 2.1.3.

2.1.2 TheThreeHundred project

In this work, we employ TheThreeHundred[2] simulation project (Cui et al., 2018). TheThreeHundred is a set of zoom-in resimulations of the Multidark Dark Matter only (MDPL2) cosmological simulation. This simulation suite extracts 324 spherical regions centered on each of the most massive clusters ($M_\mathrm{vir} > 8 \times 10^{14}h^{-1}M_\odot$) identified at $z = 0$. It 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

[1] https://www.multidark.org/
[2] https://the300-project.org/

particles are divided into dark matter and gas, following the cosmological baryonic fraction using the Planck 2015 cosmology: $\Omega_{\rm b}/\Omega_{\rm M} \approx 0.16$. This gives a combined mass resolution of $m_{\rm DM} + m_{\rm gas} = 1.5 \times 10^9 h^{-1} M_\odot$. There are 128 individual time snapshots for all 324 zoomed-in Lagrangian regions, ranging from $z = 17$ to $z = 0$.

These zoom-in re-simulations have been run with the SPH codes: 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 which incorporates full-physics galaxy formation, star formation and feedback from both SNe and AGN. In this work, we are only using the information about the simulated dark matter haloes and not the galaxies.

THETHREEHUNDRED simulations provide a useful testbed to develop the observational strategy and forecast the performance of upcoming wide-field surveys around galaxy clusters. Firstly, 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 from the cluster centre all the way to beyond $5R_{200}$ allows us to study all the environments present – from individual galaxy halos to filaments, groups, and the cluster core.

2.1.3 Halo selection

As we are interested in developing observational strategies of identifying large-scale structure, we require a tool for converting clumps of bound particles in the simulation data into galaxy haloes. The work in this thesis utilizes the widely used AMIGA Halo Finder (Gill et al., 2004; Knebe et al., 2011) to determine the halo properties. AHF operates by identifying peaks in the matter density field, and returns the position and velocity of each halo and subhalo, as well as properties such as their radii, their velocities and masses. The position of the haloes are defined by centre of mass of the particles bound to that halo. Similarly, the velocity of a halo is the average of the velocities of the particles bound to that halo. The final result is a list of halo catalogues for each of the 324 resimulated clusters, for all 128 snapshots. In this thesis, we use the information from the $z = 0$ snapshot and

only use the following columns: $X_c, Y_c, Z_c, V_{xc}, V_{yc}, V_{zc}, M_{\rm vir}, R_{\rm vir}, E, b, c, fM_{\rm hires}$, which corresponds to the x,y,z positions of the clusters in kpc, the velocities in the same planes in km/s, the virial mass and radius in solar masses and kpc, the moment of inertia tensor[3], the sphericity and asphericity[4] of the haloes and the fraction of mass in high-resolution particles.

The direct outputs of AHF place the resimulated regions with respect to the larger MDPL2 cosmological box. We briefly run through the post-processing procedure we use to appropriately center the clusters. To do this, we firstly place the positions and masses of the haloes in proper coordinates from comoving coordinates, using $h = 0.7$. Then, we place the most massive cluster halo at the origin and restrict haloes to fall within $15h^{-1}$ Mpc of the cluster main halo. Finally, as is typically done is observational studies of clusters, we put the velocities of the haloes in the reference frame of the cluster. Any further modifications of the halo catalogues we describe in the upcoming chapters.

2.2 The Multi Object Spectroscopic WEAVE Survey

WEAVE (William Herschel Telescope Enhanced Area Velocity Explorer) is a next generation multi-object-spectrograph (MOS). At the time of writing this thesis, the instrument is currently undergoing commissioning after first light data was taken in December of 2022[5]. The spectrograph makes use of $\sim$ 1000 individual fibres deployable over a 2-degree-diameter field-of-view, with both high resolution spectral observing modes ($R \sim 20000$) and low resolution modes available ($R \sim 5000$). The instrument also includes 20 small deployable integral field units (mini-IFUs), as well as one large IFU. In this thesis we are only concerned with the MOS observing mode. Further details on the instrument can be found in Balcells et al. 2010; Dalton et al. 2014; Dalton 2016; Jin et al. 2023.

The high source densities achievable from WEAVE's high fibre multiplex will

[3]the moment of inertia tensor is used to calculate the principal axes of a particle distribution.

[4]the sphericity is defined as the ratio of the third principal axis to the first principal axis. Asphericity is the ratio of the second principal axis to the first principal axis.

[5]https://www.ing.iac.es/PR/press/weave_LIFU_first_light.html.

enable a large range of scientific studies to be undertaken in the coming years. Such research areas include low redshift stellar spectra, intermediate redshift stellar populations of field galaxies and high redshift quasars. WEAVE will carry out a number of sub-surveys in the next few years, one of which is the WEAVE Wide-field Cluster Survey (WWFCS; Kuchner et al. in prep). The WWFCS is one of three WEAVE Cluster surveys, the other two being the Nearby Cluster Survey and the Cosmological Cluster survey (Jin et al., 2023). In the next section, we describe the WWFCS survey details.

2.2.1 The WEAVE Wide Field Cluster Survey

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, G. et al., 2006) and OmegaWINGS (Moretti, A. et al., 2017) surveys. The WINGS sample covers a wide range of cluster masses ($\sigma = 500 - 1200$ kms^{-1}; $\log L_{\rm X} = 43.3 - 45$ ergs^{-1}; virial masses $\log(M_{\rm cl}/M_{\odot}) = 13.8 - 15.5$). From the WINGS sample, approximately 20 clusters have been selected that are in a suitable declination range ($10 < \text{dec} < 60$ degrees), have a reasonable RA distribution and are covered by the SDSS footprint so that there are homogeneous images available for target selection. 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 prep). The WWFCS will use the low spectral resolution 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.

2.2.2 WEAVE target selection

Based off current target density estimates, the natural survey depth is $r = 19.75$, which translates to a stellar mass limit of approximately $10^9 M_\odot$ (Kuchner et al., 2020). Therefore, the target selection includes galaxies that are $r < 19.75$ and are also subject to a colour cut $-0.1 < g - r < 1.2$ which retains all normal, non-dust-reddened, cluster galaxies. To complement this, the survey will incorporate high-quality photometry from Pan-STARRS1 (PS1; Beck et al. 2020) for determination of the source redshifts as we have a complete footprint of all of the cluster regions. A photometric redshift cut of ± 0.1, (which corresponds to 2.5 standard deviations (σ)), will be applied to eliminate sources that are likely not in the vicinity of the cluster. On average, it is expected that we will have between 4000 and 6000 galaxies in each cluster structure (Jin et al., 2023), Kuchner et al. in prep.

2.2.3 WEAVE fibre allocation

An integral part of the preparation for WEAVE observations is to carry out a realistic allocation of spectroscopic fibres to the science objects, since, this process can potentially distort and limit the spatial information that can be derived from said 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. Optimising fibre allocation is not a trivial task, and sophisticated software is generally used to reduce costly human intervention. `Configure` is the program 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, as illustrated in Figure 2.1. `Configure` uses a probabilistic technique named 'simulated annealing' (Kirkpatrick S., 1983) to mimic 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 3.1). The algorithm then optimises the fibre allocation by finding the configuration with the lowest

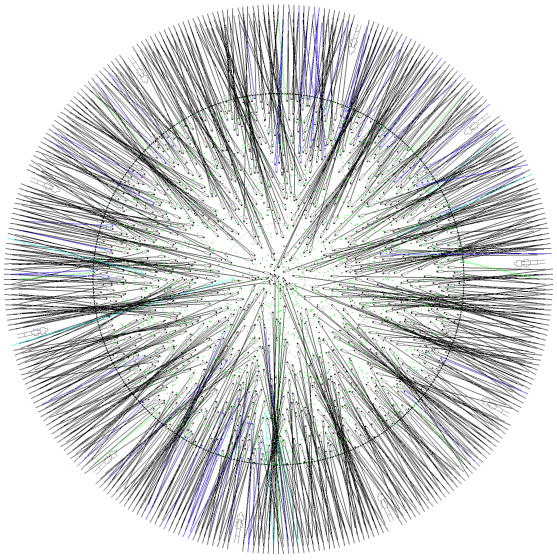


Figure 2.1: Configured field for an example WEAVE Observer Block (OB). The black outermost circle indicates WEAVE's 2-degree-diameter field of view. Each line corresponds to a fiber that has been placed on an object. For reference, black lines are science targets, cyan lines are calibration stars, blue lines are blank sky targets and green lines are guide stars. This is Figure 2 in Jin et al. (2023).

'energy' by swapping the position of fibres until the minimum is found.

This process determines the objects that will be allocated a fibre and therefore decides which targets will have spectroscopic information. In our case, 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 Chapter 3.

2.3 Topological structures extractor

A key element of this thesis is the extraction of cosmic web components. In this section, we describe the main tool we use to do so in our work.

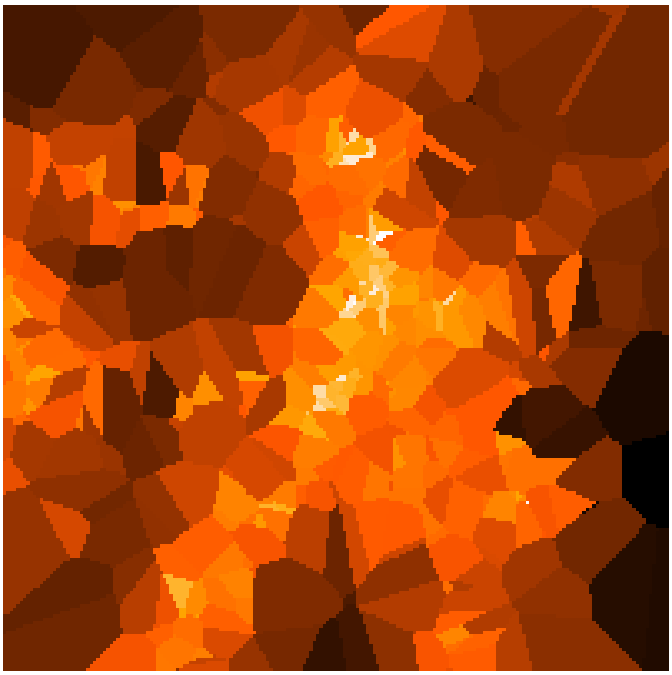


Figure 2.2: Visualization of the Delaunay tessellation for a simulated cluster from TheThreeHundred project. The size of the cell corresponds to the inverse of the density whilst the brightness corresponds to the density.

2.3.1 DisPerSE

To map the large-scale structure around clusters and extract filaments, we utilize the topological structures extractor `DisPerSE` (Sousbie, 2011; Sousbie et al., 2011), which uses the concept of Morse theory (Stein et al., 1963) and the theory of persistence applied to matter distributions. The software takes, as input, a discrete set of points in 2D or 3D to reconstruct the volume as cells, faces, edges and vertices. The Delaunay Tessellation Field Estimator (DFTE; Schaap & van de Weygaert 2000; Cautun & van de Weygaert 2019) calculates the density field by performing a Delaunay tessellation: it divides the space into triangles (tetrahedra in 3D) whose vertices are formed by the point distribution. The size of the triangles (tetrahedra) is a measure of the local density and is defined at the location of each sampling point by the inverse of the area (volume) of its surrounding Delaunay triangles (tetrahedra). The density estimates are then interpolated to any other point by assuming that the density inside each triangle (tetrahedra) varies linearly. To construct the filamentary network from the density field, `DisPerSE` then extracts the critical points, (points where the gradient of the density

field is zero), such as maxima, minima and saddle points. Nodes are identified as the maxima. Arcs linking maxima to saddle-points trace the filamentary structures. The output of this algorithm is a list of segments that combine to form a filamentary skeleton that trace the topologically significant regions in the density field.

2.3.2 Input parameters

When running `DisPerSE`, there are two main user chosen input parameters: the *persistence* and the *smoothing*. Further to this, there is an optional argument to specify how the boundary conditions are dealt with. The last option is to use mass-weighting in the structures extraction. These features are defined as the following:

1. *Persistence*: Persistence is the ratio of the density value between a topologically significant pair of critical points (Kraljic et al., 2022). Expressed in terms of numbers of σ, persistence quantifies the significance of the critical pairs in the Delaunay tessellation of a random discrete Poisson distribution and is analogous to the signal-to-noise ratio often used in observational astronomy. Figure 2.3 shows a filament network plotted with two differing persistence thresholds. The top row uses a persistence of 2σ and the bottom row uses a persistence of 3σ. It is evident there are features in the lower persistence network that do not appear in the higher persistence network as a result of filtering noise and removing less significant filaments. 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.

2. *Smoothing*: Smoothing is the second main input parameter that dictates the rigidity of a filament network. Smoothing is achieved by averaging the position of each vertex with that of its direct neighbors. A low level of smoothing means the filament paths are not forced into straight lines. A higher smoothing level results in much more rigid filament paths.

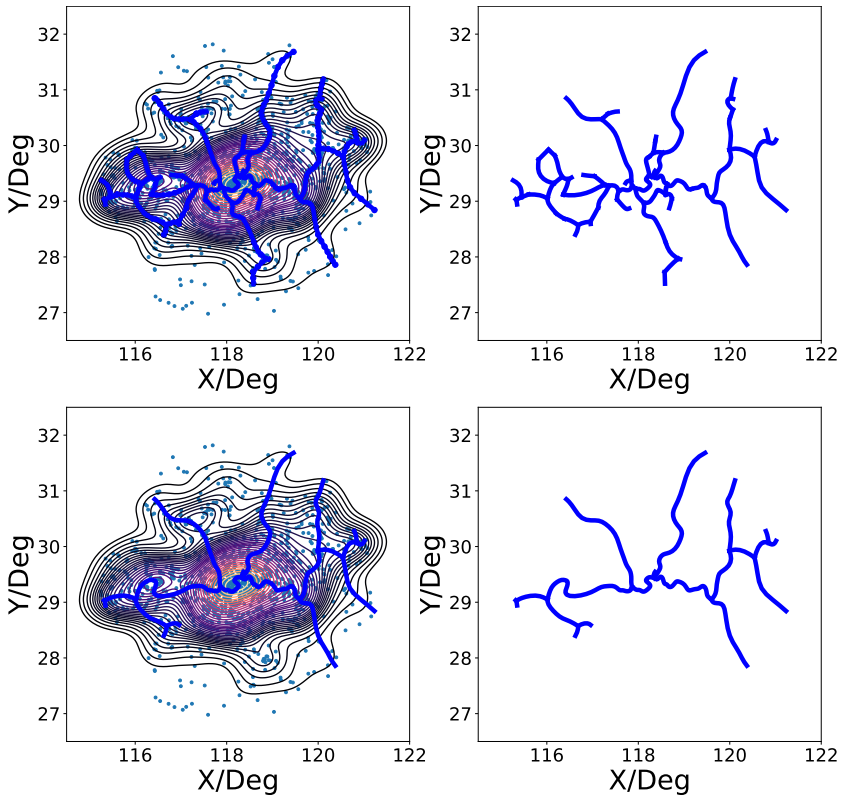


Figure 2.3: Example filament network extracted in 2D with different persistence thresholds. Top row: persistence = 2σ. Bottom row: persistence = 3σ. All plots correspond to the same cluster. Left column: KDE smoothed matter density distribution with the extracted filament network overlaid. Right column: filament network only. This is taken from Figure 5 in Cornwell et al. (2022).

3. *Boundary conditions*: Boundary conditions are the criteria that we apply to deal with the apparent physical boundaries in the input particle distributions. A careful consideration of the appropriate treatment of the boundary is needed. Given that we are extracting topological structures over a finite volume, performing a hard cut on the input particles at the bounding box causes the generation of spurious filaments near the boundary. To account for this, `DisPerSE` includes an optional argument to deal with the boundaries. These options include a mirrored distribution, a periodic boundary condition, a smoothed boundary condition (a surface of guard particles outside of the bounding box are added by interpolating the estimated density computed from the distribution inside the bounding box) and a void boundary condition (cf. Sousbie, 2011). We discuss the use of this argument further in Section 3.3.5.1.

4. *Mass-weighting*: mass-weighting associates to each vertex of the tessellation a weight corresponding to the mass of the halo at that vertex. Using mass weighting requires an adaption of the input persistence, such that we require a higher persistence to recover a similar looking network compared to a non-mass weighted network. This is due to the higher contrast in density computed during the tessellation due to the large halo mass range used in TheThreeHundred. The choice of using mass-weighting 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; Laigle et al., 2017; Kraljic et al., 2017; Sarron et al., 2019). Furthermore, Kuchner et al. (2020) showed that mass-weighting the filamentary networks makes the filamentary networks more reliable.

Chapter 3

Filamentary networks in galaxy cluster simulations

A large fraction of the work completed in this thesis revolves around the extraction of cosmological structures around galaxy clusters. One of the final steps in taking hydrodynamical simulations to fully-informative mock observations is allocating a realistic fibre configuration. Only then can we understand the success and limitations in our ability to reconstruct filamentary networks around galaxy clusters. This is an essential test as without it, we will be unsure as to how well the structures we extract in future surveys will represent the 'ground truth'. I.e., by taking out this work, we can question whether the filament networks we extract with the WWFCS are representative of those existing in the real Universe. In this chapter, we investigate this by using TheThreeHundred galaxy cluster simulations to design WEAVE-like mock observations. After incorporating the observational selection effects, such as the projection, fibre configuration and application of magnitude limits, we quantify the difference in the large-scale structure we trace in the simulations compared to our 'mock observations'. Future wide-field spectroscopic observations of galaxy clusters will target a high percentage of galaxies. Furthermore, we show that these surveys possess the ability to accurately trace cosmic filaments and reconstruct global parameters of the networks, such as the cosmic connectivity. Finally, we discuss the outlook of this work with regards to these upcoming observations. The content in this chapter was published in Cornwell et al. (2022).

3.1 Introduction

For the past few decades, research has focused on galaxy clusters when studying the interplay between environment and galaxy evolution (Gray et al., 2009; Balogh et al., 2017). There is a 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 intra-cluster medium of cluster cores during their infall. Ram pressure stripping is one of several possible mechanisms quenching galaxies infalling onto 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 (Walker & Lau, 2022; Galárraga-Espinosa et al., 2022; Gouin et al., 2022).

Galaxy clusters are therefore not isolated islands, but assemble, replenish and grow via ongoing mergers with smaller clusters, 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; Salerno et al., 2020; Gouin et al., 2020; Malavasi et al., 2022). However, they are challenging to capture.

Whilst filaments can be identified by mapping the gas distribution of galaxy clusters in simulations (Kuchner et al., 2020; Vallés-Pérez et al., 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 Two-Degree Field Galaxy Redshift Survey (2dFGRS, Colless et al. 2001), the Galaxy and Mass Assembly survey (GAMA, Liske et al. 2015), the VIMOS Public Extragalactic Redshift Survey (VIPERS, Mohammad et al. 2018), 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., OmegaWINGS Gullieuszik et al. 2015; Moretti, A. et al. 2017 or GOGREEN Balogh 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 prep) and the 4MOST CHileAN Cluster galaxy Evolution Survey (CHANCES; Haines et al., 2023). We motivate our studies with the WWFCS, which will cover 20 low redshift ($z \sim 0.05$) galaxy clusters in a mass range of $\log(M_{\rm cl}/M_{\odot}) = 13.8 - 15.5$ 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, (Kuch-

ner et al., 2021), careful preparation is required. We have planned, tested, and fine-tuned our steps to map and characterise the infall regions of these clusters with a large statistical sample of simulated clusters from TheThreeHundred survey, and investigated strategies for doing so in redshift space (Kuchner et al., 2020, 2021; Kuchner et al., 2022; 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 chapter 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 fibres 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 Chapter, 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.

This chapter is structured in the following way: Section 3.2 describes the data we have used. This includes the spectroscopic survey inspiring this chapter and the numerical simulations used to create mock observations. Section 3.3 reports our generation of mock observations and a summary on how we extract the cosmic web. Section 3.4 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 3.5.

3.2 Galaxy cluster information

The framework presented in this chapter 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.

3.2.1 The WEAVE Wide-Field Cluster Survey

As described in Chapter 2.2, the WWFCS will obtain thousands of galaxy spectra per cluster for up to 20 low redshift galaxy clusters ($0.04 < z < 0.07$). Careful consideration of the observing strategy is required for each individual cluster. Due to variations in their masses and redshifts, the resulting angular diameter distance of the cluster region will vary from cluster to cluster. An example WWFCS observing strategy is illustrated in Figure 3.1. We show a simulated galaxy cluster (cf. Section 3.2.2) 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 under-dense 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.

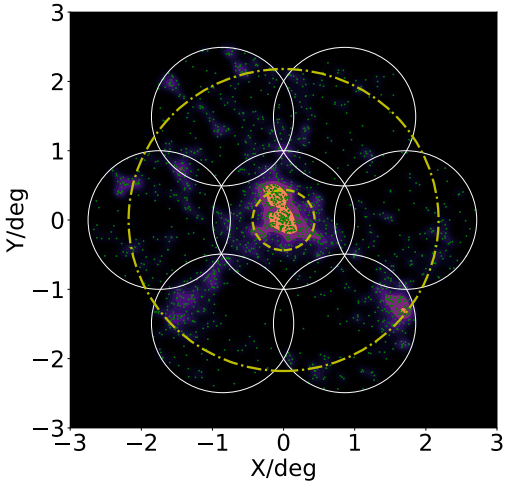


Figure 3.1: Example simulated cluster from TheThreeHundred with a similar mass and redshift to the cluster Abell 602, one of the WWFCS targets. The projected dark matter density distribution is shown, derived using a Kernel Density Estimation (KDE) with a 500 kpc smoothing scale. This box has a depth of 10 Mpc. The green dots indicate the positions of galaxy-mass halos. Each white circle encloses a WEAVE field with a 2 degree diameter. The central yellow dashed circle corresponds to R_{200} and the larger dot-dashed yellow circle to $5R_{200}$.

3.2.2 TheThreeHundred galaxy cluster simulations

In this work, we use TheThreeHundred project of galaxy cluster resimulations, as described in detail in Section 2.1.2. This dataset has already been used previously to generate theoretical expectations with the WWFCS in mind. For instance, Kuchner et al. (2021) 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. (2021) forecast that the identification of the cosmic 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.

3.3 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.

3.3.1 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 maximising the observational efficiency. The WWFCS can place MOS fibres on targets over a 2-degree-diameter field (Figure 3.1). In order to map the filamentary structures that feed clusters it is necessary to maximise 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 prep.). 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.1 (see examples in Figures 3.1, 3.3, and 3.4). The 'simul-observations' described below follow the same strategy. We note that by applying this optimisation process we have been able to reduce the required number of pointings (and thus the required observing time) by $\sim 15\%$ 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.

3.3.2 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} of the clusters using their spectroscopic redshift z and velocity dispersion σ from the WINGS survey (Moretti, A. et al., 2017) using the following equation (Poggianti et al., 2010; Finn et al., 2005):

$$R_{200} = \frac{1.73\sigma}{1000\text{km s}^{-1}} \frac{1}{(\Omega_\Lambda + \Omega_\text{M}(1+z)^3)^{1/2}}. \tag{3.1}$$

Here, R_{200} is measured in Mpcs and σ is measured in km/s. 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 prep.). The bottom panel of Figure 3.2 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 2.2). We address this mismatch in next section.

3.3.3 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 halos 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 centered 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

($fM_{\rm hires} > 0.99$). 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 halos with a mass greater than the simulation resolution ($3 \times 10^{10} h^{-1} M_{\odot}$), corresponding to 20 dark matter particles, as explained in Kuchner et al. 2020.

3.3.3.1 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 halos evolve self similarly (Kaiser, 1986; Mostoghiu et al., 2018). Self-similarity implies that the dark matter distribution (and hence the location of dark-matter halos) 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 halos 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 A.2. 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 $M_{\rm F}$ 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 halos in each simulated cluster approximately matches that expected in the the observed cluster. For each WWFCS cluster, ten analogue clusters with similar masses are randomly selected from the scaled clusters.

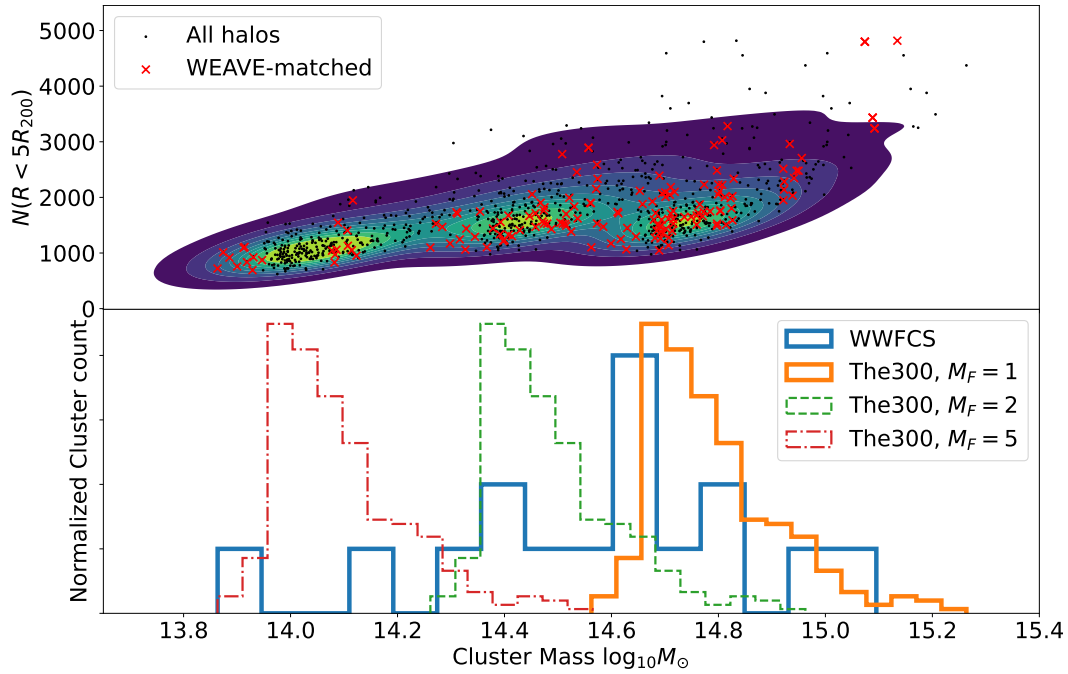


Figure 3.2: Bottom panel: normalized cluster count as a function of cluster mass showing the mass distribution of the WWFCS clusters (solid blue line) and TheThreeHundred un-scaled clusters ($M_{\rm F} = 1$, solid orange line), and scaled by $M_{\rm F} = 2$ (green dashed line) and $M_{\rm F} = 5$ (red dot-dashed line). Three scaling values ($M_{\rm F} = 1, 2$, and 5) are sufficient to fully cover the mass range of the WWFCS cluster sample. Top panel: the number of galaxy halos contained within $5R_{200,\rm scaled}$ 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.

Figure 3.2 demonstrates that three values of the mass-scaling factor $M_{\rm F}$ (1, 2 and 5) are sufficient to provide enough mass-matched simulated clusters. Of course, $M_{\rm F} = 1$ implies no scaling is applied. The bottom panel shows the mass distribution of the WWFCS clusters (black) 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.

3.3.3.2 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 $R_{200,\mathrm{scaled}}$ can be derived from the relationship between R_{200} and M_{200}, adapted from Poggianti et al. 2010,

$$R_{200}^3 = \frac{M_{200}}{K \times h^2(\Omega_\Lambda + (1+z)^3\Omega_0)}, \tag{3.2}$$

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

For each WWFCS cluster the angular diameter distance is calculated using their redshifts (Moretti, A. et al., 2017) with the adopted cosmology. Next, for each of the corresponding ten 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 Figure 3.2 displays the number of galaxy halos contained within $5R_{200,\mathrm{scaled}}$ from their cluster centers 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 prep.; see also Figure 3.1 and Appendix A.1), and therefore the number of potential targets for each cluster shown in the figure is a conservative lower limit.

3.3.3.3 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.1. Only halos covered by these

pointings will be considered as possible spectroscopic targets.

Galaxy-size dark-matter halos 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 chapter, but the fibre allocation program `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

$$r_{\rm total} = W - 2.5\log_{10}(M_{\rm halo}/M_{\odot}), \tag{3.3}$$

where W is a constant that is calculated by mapping the least massive halos in each simulated cluster (Appendix A.2) to the planned r-band limit of the WWFCS spectroscopic observation ($r_{\rm total} < 19.75$, corresponding to an approximate galaxy stellar mass limit of $\sim 10^9\,M_{\odot}$, Kuchner et al., in prep). 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 $r_{\rm fibre} < 21.50$, which sets the planned exposure times of ~ 1 hour. This exposure time is expected to yield reasonable signal-to-noise ($S/N > 5$ per Å, for all the spectra up to this magnitude limit), and we therefore expect close to 100% redshift completeness for the observed (and thus mock) galaxies (see Kuchner et al. in prep.).We use a simple procedure to allocate magnitudes to the galaxy-sized dark matter halos here, as the results in this chapter 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 Figure 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 which are not necessary for our purposes.

At this point we have created a set of 160 simulated galaxy clusters (10 per

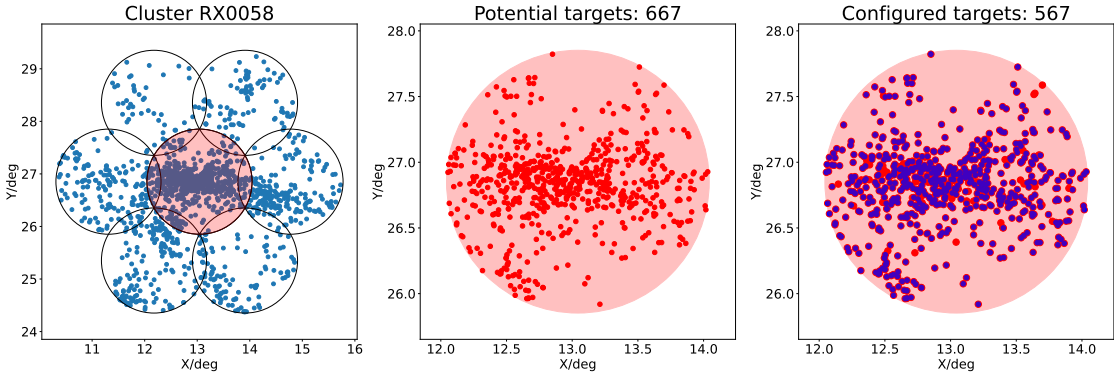


Figure 3.3: A demonstration of the process of 'configuration', the allocation of spectroscopic fibres to targets, on a simulated cluster mass-matched to WWFCS cluster RX0058 ($M \sim 4.3 \times 10^{14} M_{\odot}$, $R_{200} \sim 1.54\,\mathrm{Mpc}$), an average mass WWFCS cluster. Left panel: mock 'observation' of the simulated cluster containing 8 2-degree 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 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.

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.

3.3.4 Allocating spectroscopic fibres to mock galaxies using `Configure`

To make a more realistic analogue to the planned observations, before running `Configure` (see Section 2.2.3 for details) we pollute our target catalogues with background objects. We randomly place 1400 background objects on each WEAVE field, corresponding to a number density $\sim 450\,\mathrm{deg}^{-2}$. This is somewhat larger than the galaxy number density corresponding to an in-fibre magnitude limit $r_{\mathrm{fibre}} < 21.5$, the planned WWFCS limit. 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 haven't been assigned to cluster members according to their photometric redshifts). This is especially

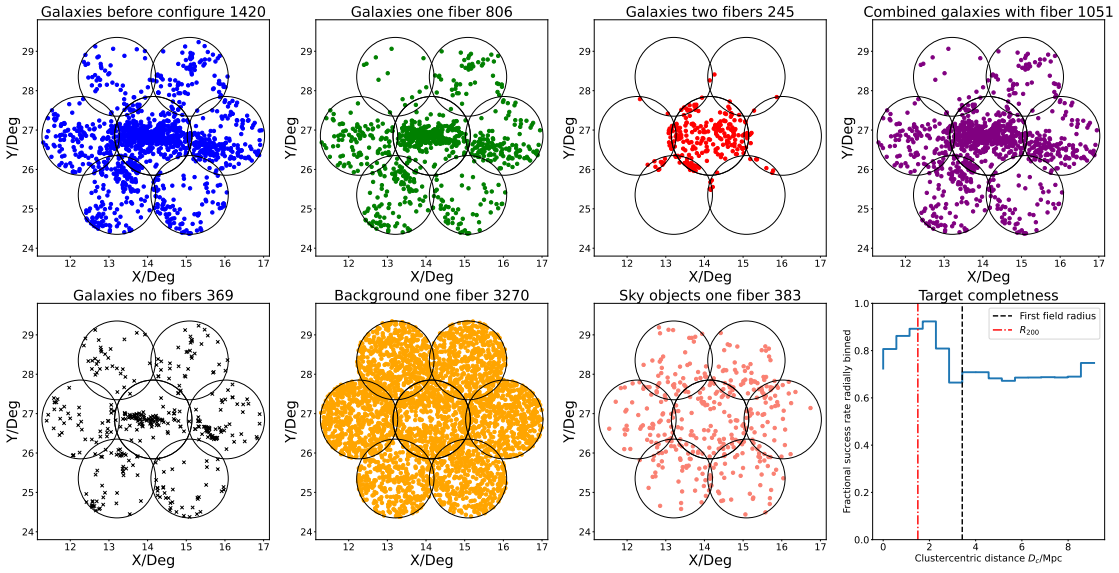


Figure 3.4: Results for the full configuration process of one simulated analogue to cluster RX0058. We plan to observe this cluster using 8 WEAVE individual pointings (black circles). The central pointing will be observed twice given the high density of targets in the cluster core. For this cluster, 74% of its simulated galaxies have at least one fibre allocated and 69% 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 3.4.1) in clustercentric distance bins. R_{200} and the radius of the central WEAVE field are plotted for reference. The bin width is 1/6 of of the radius of one WEAVE pointing (i.e., 1/6 of one degree).

relevant further away from the cluster center as fibres are then free to be placed on background targets if they are brighter than the magnitude limit. In this chapter, we stress-test this assumption by putting a slightly exaggerated number of background targets in the catalog 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 3.3.5.1. For this chapter, 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 `TheThreeHundered` volume. This is supported by our tests using available observational data of the cluster centers. 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-prioritised. We keep them in the target catalog for the sole reason that unallocated fibres can be used and they do not feature in our filament finding, as described in Section 3.3.5.1.

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., in prep.

Therefore, each WEAVE field target list consists of $N_{\rm gal}$ cluster galaxies, determined by the simulated cluster galaxy sample (see Section 3.3.3.3 and Figure 3.2), $N_{\rm sky}$ 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 pointing overlap (Figures 3.3 and 3.4) 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 prep.). The aim of the process is to maximise the number of target galaxies with at least one fibre allocated. 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 3 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 $\sim$ 10–20% 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 fibres. 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 3.1)[1].

For all simulated clusters we sequentially apply `Configure` to each WEAVE field (see Figure 3.4). 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 3.1). The process is illustrated with one example for the cluster RX0058 in Figures 3.3 and 3.4.

Obviously we are not able to allocate a fibre to each target cluster galaxy. In the typical cluster shown in these figures, 74% of the cluster galaxies have at least one

[1]In this exercise we do not include the flux calibration targets and guide stars since given their small numbers they have a negligible effect in our results.

Table 3.1: 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

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. We note that a high density of background targets may impact the completeness of our target galaxies in the outskirts of clusters, despite the different target priorities we have applied. Furthermore, for each WEAVE field, the target completeness varies spatially due to fibre placement issues (Hughes et al., 2022). In the next section, we will analyse quantitatively the effect the configuration process has on our ability to map the large-scale structure and filaments around galaxy clusters.

3.3.5 Cosmic web extraction method

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

3.3.5.1 DisPerSE

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, 2021). A full description of the structures extraction tool `DisPerSE` is provided in Chapter 2.3.

In the analysis that follows we will use a persistence of 2.5σ for the simulated ref-

erence 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.

As described in Chapter 2.3, careful treatment of the boundaries is required. Whilst it has been shown that the smoothed boundary condition can be used successfully for observational catalogues confined to limited volumes, it is not sufficient for our WEAVE pointings. The complex 2D geometrical shape defined by the positions of the WEAVE pointings that will tile each cluster and its environment (Appendix A.1) 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 covered by the planned WEAVE pointings with a random uniform distribution of artificial galaxies that will act as 'guard particles' 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.

3.4 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 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.

3.4.1 Recovery of cluster galaxies after `Configure`

Physical constraints from the fibre positioner imply that we will never have a 100% 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 $72.7\% \pm 1.7\%$, 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 $81.7\% \pm 1.3\%$. We argue that this value is more relevant than the overall one when dealing with the characterisation of the filament network since the whole cluster core will behave just as a single node (Kuchner et al., 2020, 2021).

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 $\sim 77\%$, while the corresponding fraction of low-mass galaxies is $\sim 69\%$. 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 $\sim 2R_{200}$ the fractions are approximately equal.

We find that, whilst 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 (Figure 3.4), 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 prep.).

3.4.2 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.

3.4.2.1 Skeleton distance

A useful metric designed to quantify the accuracy of filament extraction is the 'skeleton distance' D_{skel} (Laigle et al., 2017; Malavasi et al., 2016; 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 Figure 3.5. The left 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 panel, where the differences between the red and green networks are largest, showing the individual segments. The right panel illustrates how we calculate cluster connectivity and is discussed in 3.4.2.2. The $D_{\rm skel}$ 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 $D_{\rm skel}$ values will not be necessarily the same (see below). In both cases, the distribution of $D_{\rm skel}$ values quantifies how well both filament networks match each other.

Figure 3.6 shows the reference network (left 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 panel shows the normalised probability distribution function for $D_{\rm skel}$, 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 ($\sim 1\,$Mpc; Kuchner et al. 2020). A large proportion of $D_{\rm skel} > 1$ Mpc would indicate that a filament in this cluster has no counterpart in the corresponding mock-observational cluster. Note that, typically, the median $D_{\rm skel,R2C}$ is smaller than the median $D_{\rm skel,C2R}$ 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 $D_{\rm skel}$ median values will not only be very small, but also very similar to each other. Therefore, the median values of $D_{\rm skel}$ 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 A.3).

Another measure of the similarity between the reference and 'configured' filament networks is provided by the fraction of $D_{\rm skel}$ values that are larger than $\sim 1\,$Mpc (the typical radius of filaments). The right panel of Figure 3.6 shows that this

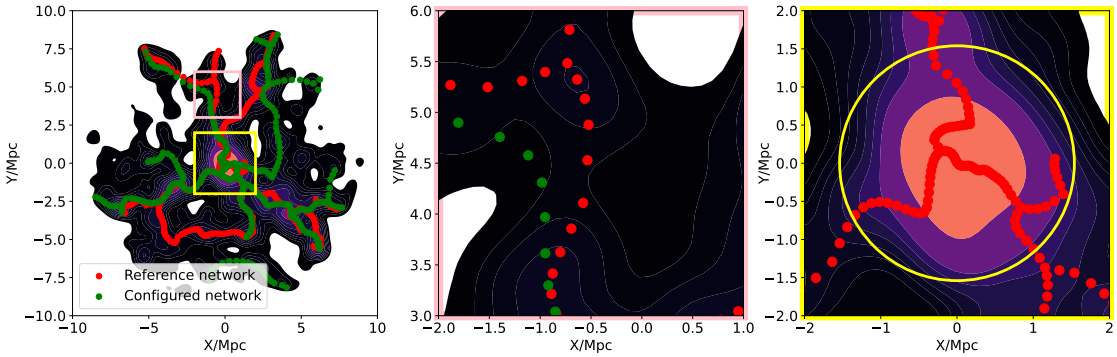


Figure 3.5: Illustration of the methods used to calculate $D_{\rm skel}$ and cosmic connectivity. Left panel: the reference filament network (red) and 'configured' network (green) are plotted on top of the KDE-smoothed halo density distribution of a simulated cluster, an analogue to WWFCS cluster RX0058. 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 panel, where the two networks show large differences to demonstrate how $D_{\rm skel}$ is calculated (see Section 3.4.2.1). Right panel: zoom-in on the cluster core (yellow box in the left 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 3.4.2.2). For this cluster, the connectivity has a value of three.

fraction is also reassuringly small ($\sim 10\%$) in both cases.

3.4.2.2 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 Figure 3.5 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 (Sarron et al., 2019; Darragh Ford et al., 2019; Gouin et al., 2021), albeit with considerable scatter. Our simulated clusters show a similar correlation. If the recovered filament network is similar to

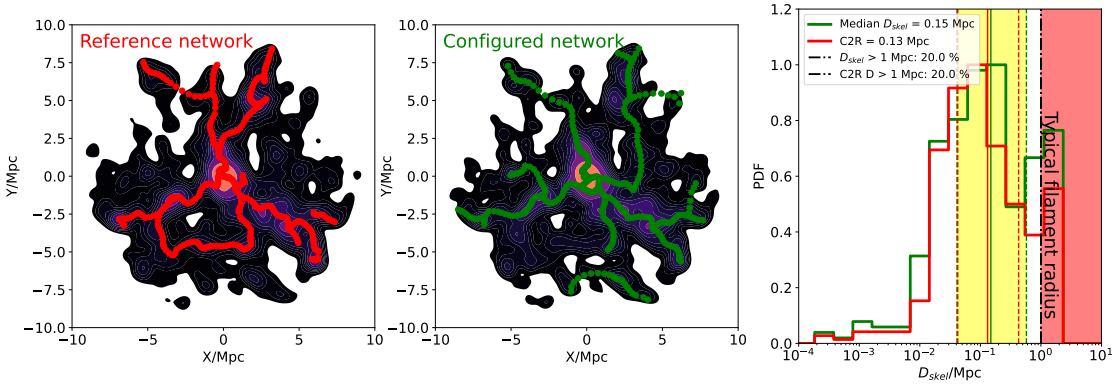


Figure 3.6: Illustration of the recovery of the filament network around a simulated analogue of the RX0058 cluster, the same cluster as Figure 3.5. Left panel: KDE-smoothed density distribution of the simulated cluster galaxies with reference filament network in red. Middle panel: as the left panel, but showing the density distribution and filament network (green line) recovered from the 'configured' (mock-observed) galaxy sample. Right panel: $D_{\rm skel}$ 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 correspond to the 25th and 75th percentiles. The dot-dashed black line represents the typical radius of a filament ($\sim 1\,$Mpc). The values of the medians and the percentage of segments with $D_{\rm skel} > 1\,$Mpc are shown in the legend. We normalized the PDF's to have a common peak value.

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.

3.4.3 Quantifying the quality of the recovered filament networks

We are now in a position to use the $D_{\rm skel}$ 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, Figure 3.6 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% 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 $D_{\rm skel}$ are 0.12 Mpc going in the R2C direction and 0.16 Mpc going in the C2R direction (cf. Section 3.4.2.1). These values are much smaller than ~ 1 Mpc, the typical radius of filaments. Moreover, only 8% and 10% 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 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.

Figure 3.7 confirms that the $D_{\rm skel}$ results shown for the RX0058 analogue are indeed typical of the whole sample. We can therefore stack the $D_{\rm skel}$ distributions for the 160 simulated clusters (Figure 3.8) and derive representative average quantities for the whole sample. On average, the median values of $D_{\rm skel,R2C}$ and

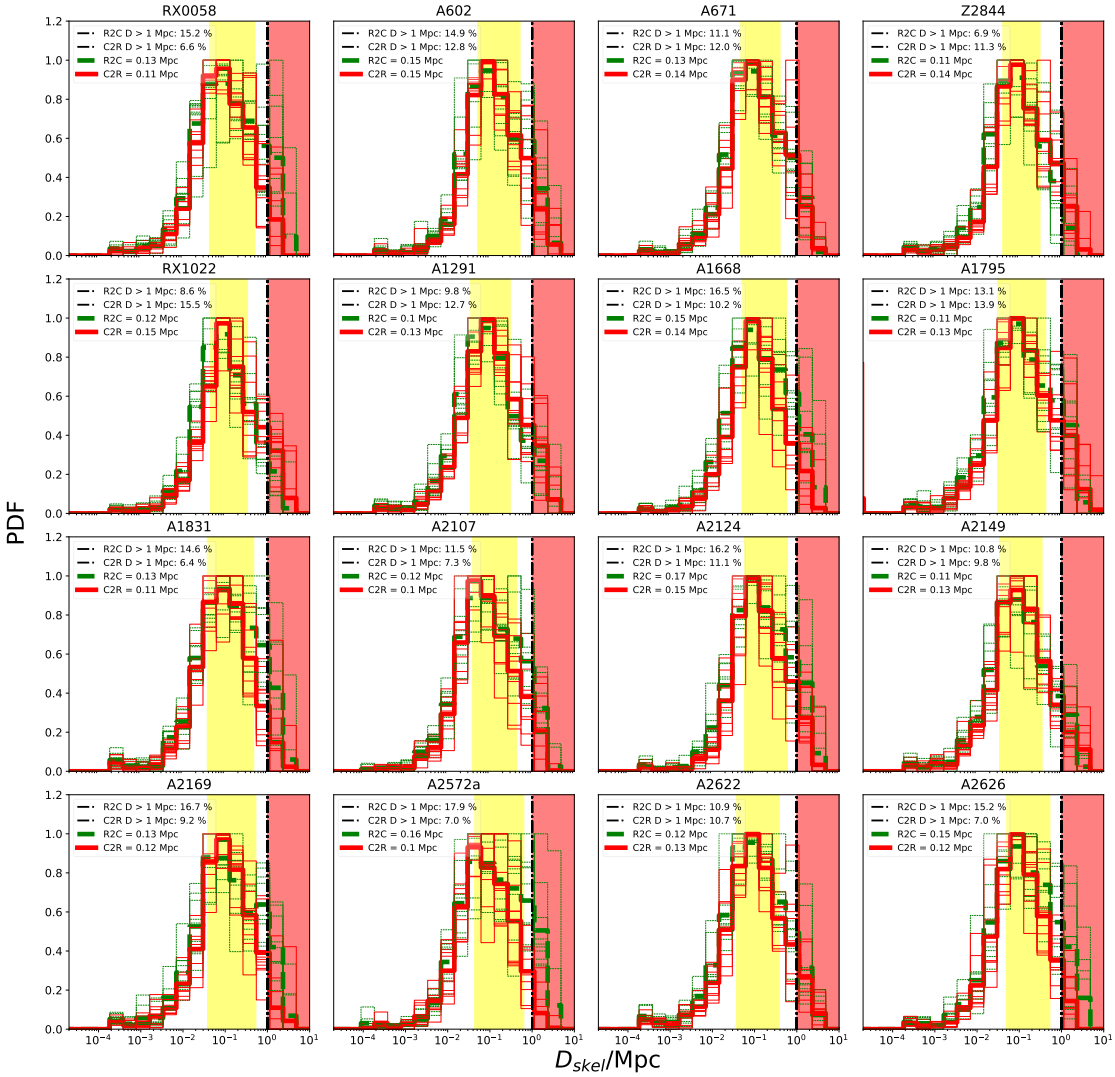


Figure 3.7: $D_{\rm skel}$ distributions for all the simulated WWFCS cluster analogues. Each panel shows the individual cluster comparison (thin lines) and the average for the 10 simulated cluster mass-matched to each WWFCS cluster (thick lines). The format of each panel follows that of the right-hand panel of Figure 3.6. There is little variation in $D_{\rm skel}$ over different WWFCS analogue clusters and the $D_{\rm skel}$ median is always much less than a typical filament radius of 1Mpc.

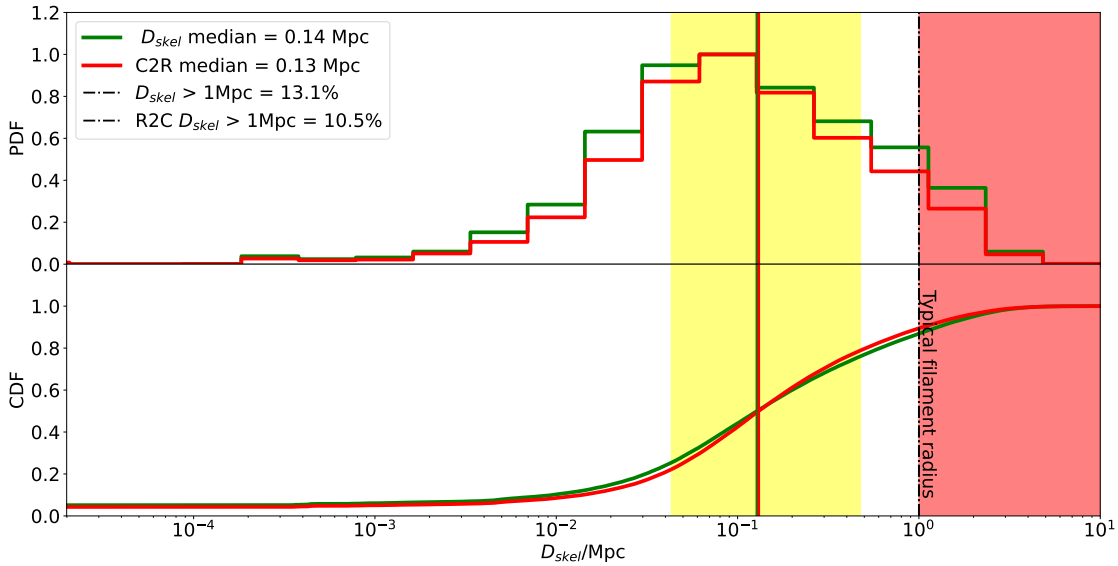


Figure 3.8: Stacked probability density function (top panel) and cumulative density function (lower panel) of the $D_{\rm skel}$ distributions corresponding to all the simulated clusters shown in Figure 3.7. The positional difference in the networks is minimal – our mock observations of filaments around galaxy clusters successfully recreate the 'true' simulated filament network.

$D_{\rm skel,C2R}$ 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% of the corresponding filamentary segments are more that 1 Mpc away from each other.

Figure 3.9 shows a comparison between the connectivity of the reference filament networks $C_{\rm reference}$ and that of the 'configured' ones $C_{\rm configured}$. 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.

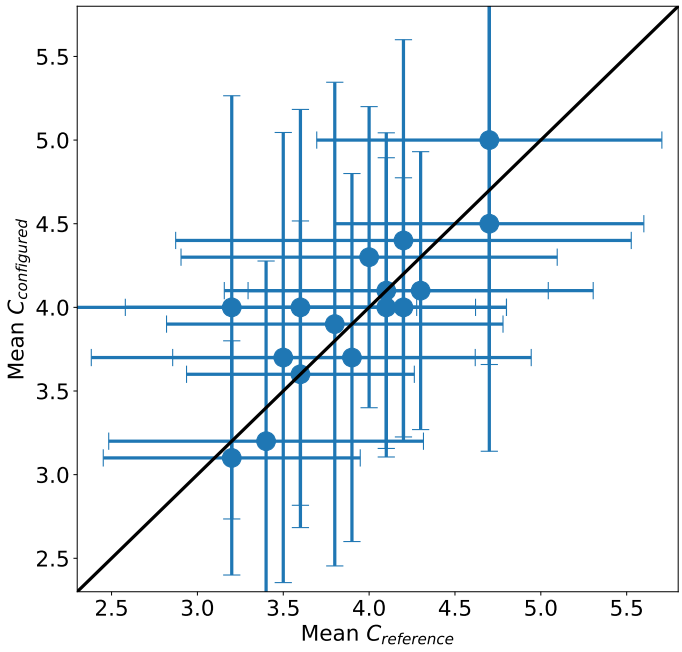


Figure 3.9: Comparison between the cluster connectivity C of the reference and 'configured' filament networks. The 1-to-1 line is shown in black. Each point corresponds to the mean connectivity of the 10 mass-matched analogues of each WWFCS cluster, and the error bars show the $\pm 1\sigma$ scatter.

3.5 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 Weave Wide Field Cluster Survey (WWFCS[2]) will explore and map in detail these complex regions, in this chapter 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 prep.). 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 summarise our main results below.

1. 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 $72.7\% \pm 1.7\%$ 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 – 4062. The high completeness that the WEAVE observations will allow, together with the large field coverage, are key to the success of the survey.

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

2. 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 $D_{\rm skel}$ in the reference and recovered networks is only $0.13 \pm 0.02\,$Mpc on average, an much smaller than the typical filament radius of $\sim 1\,$Mpc. Furthermore, only $\sim 11 - 13\%$ of all recovered filament segments lie at a distance larger than $1\,$Mpc away from their corresponding reference segment.

3. 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, 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.

3.5.1 Caveats and potential tests

Upon reflection of this work, we have noticed a potential bias in the methodology we have employed for calculating $D_{\rm skel}$. The right panel of Figure 3.5 illustrates

that there are a higher sampling of filament segments closer to the cluster core, potentially due to the large densities that are found in this region. Furthermore, in the core, the maximum separation of filamentary segments between a "mock observational" network and a "simulated" network is capped at roughly 1 - 2 Mpc. As a result, a large proportion of our $D_{\rm skel}$ distributions that lie at $D_{\rm skel} < 1$ Mpc will correspond to segments inside the cluster core. This implies that our $D_{\rm skel}$ metric is biased and not truly representative of the accuracy of filament reconstruction. An important potential test would be to calculate $D_{\rm skel}$ by omitting the segments within R_{200}. This would provide a better test of the reconstruction of filaments in the outskirts of clusters. We conclude that our $D_{\rm skel}$ distributions would likely be more heavily skewed towards larger distances. However, this change would not effect the calculation of the connectivity.

Chapter 4

Allocating galaxies to cosmic web environments

The mass content and morphology of different cosmic web structures, such as clusters, filaments, groups and voids are distinctly different. The influence of the environment of lower density contrast structures, such as cosmic filaments, on galaxy evolution is currently a topic of debate. In Chapter 3, we established a framework for extracting large-scale cosmic filaments around galaxy clusters after designing mock observations. A goal of future surveys such as the WWFCS is to understand the role that different cosmic web environments play on galaxy evolution. However, to be able to probe this relationship, we need to understand how accurately we can allocate individual galaxies to different environments in the first place. As we have shown that the WWFCS will accurately reconstruct filamentary networks around galaxy clusters, we now test the environmental classification of individual galaxies using TheThreeHundred project. Especially in close proximity to the core of a galaxy cluster, allocating galaxies to cosmic web filaments in observations is highly uncertain. Despite this, crucially, our success rates outperform a random classification. This means that with large spectroscopic samples of galaxies that the WWFCS will provide along with the correct statistical treatment we provide in this chapter, we will be able to draw robust conclusions on the relationship between galaxy properties and their environment. The content of this Chapter was published in Cornwell et al. (2023a).

4.1 Introduction

Galaxy clusters are an intrinsic part of the cosmic web. As a result, along with the galaxies inhabiting them, they co-evolve with the cosmic web. 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, A. L. B. et al., 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 emphasised by the fact that, even though filaments and clusters only contain 6% and 0.1% respectively of the volume of the present-day universe, they harbour 50% and 11% 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 et al., 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., 2020). 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 prep) and the 4MOST CHileAN Cluster galaxy Evolution Survey (CHANCES; Haines et al., 2023). This chapter focuses on the first of these survey. To fully exploit the extensive data generated by the WWFCS, we are developing specific analysis techniques to test and optimize the detection and characterisation of the filamentary networks around the clusters (Kuchner et al., 2020, 2021; Kuchner et al., 2022; Cornwell et al., 2022) using simulated clusters from TheThree-Hundred project (Cui et al., 2018, 2022). 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 $\sim 10\,\mathrm{Mpc}$. Such accuracy cannot be achieved with photometric redshifts.

Bringing the TheThreeHundred simulations one step closer to observations, we showed in Chapter 3 that filaments can be successfully extracted from datasets similar to those expected from the WWFCS. In this chapter, 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 chapter is structured as follows. Section 4.2 describes the (expected) observational and simulated datasets. Section 4.3 presents the reference filament networks derived from the simulated galaxy samples and defines the 'true' environment of each galaxy. Section 4.4 describes how the filaments are found in the mock WWFCS-like observations, and how the 'observed' environment of each

galaxy is determined. Section 4.5 compares the 'true' and 'observed' environments of the galaxies, and provides the necessary statistics to quantify the success (or otherwise) of the comparison. Section 4.6 summarises the conclusions.

4.2 Dataset

4.2.1 WEAVE Wide Field Cluster Survey

We describe the observational survey motivating this work in Section 2.2. During our description of the observational strategy in Chapter 3, we stated that, 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. WEAVE uses an algorithm named `Configure` (Terrett et al., 2014) to optimise fibre allocation. In Chapter 3, we showed that `Configure` is able to allocate fibres to $\sim 75\%$ of the target galaxies overall, reaching 81% outside R_{200} and that this is enough to reliably reconstruct the filament networks around the clusters. In the current chapter, we base our analysis on mock-observed (i.e., "configured") simulated galaxy samples in clusters and their outskirts.

4.2.2 TheThreeHundred simulations of galaxy clusters

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 Chapter 3 we mass-matched each of the 16 main clusters targeted by the WWFCS with 10 clusters in `TheThree-Hundred` simulations, and mock WWFCS observation were created for all 160 of them. 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). Going forward, we refer to these cluster galaxy samples as the 2D mock-observations.

Throughout this chapter 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.

4.3 Defining cosmic web environments

4.3.1 Cosmic web extraction

Following our work in Chapter 3, we employ the widely used structures extractor algorithm `DisPerSE` (Sousbie, 2011; Sousbie et al., 2011) to identify filaments in the simulation boxes at $z = 0$, which is explained in detail in 2.3. 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 $10^{10} M_{\odot}$, approximately corresponding to stellar masses larger than $10^{9}\, M_{\odot}$ [1]. Furthermore, following Kuchner et al. (2020) and following our description in Section 2.3, in this work we incorporate mass-weighting.

4.3.1.1 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 in Chapter 3, we use a *persistence* value of 4.6σ and a *smoothing* of 5. 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

[1] We note that the halo and stellar mass limits are slightly higher for more massive clusters, as described in Chapter 3

second parameter, *smoothing*, creates smoother networks by averaging the position of each vertex with that of its direct neighbors.

4.3.1.2 2D mock-observational filament networks

We also need to define filament networks using a mock-observational dataset similar to the one WEAVE will provide. The main two differences from the original simulations are that we will not observe every single galaxy (see Section 4.2.1) and that we will not have radial distances but only redshifts (or radial velocities). As explained in Chapter 3, despite these limitations, reliable 2D filamentary networks (projected on the plane of the sky) can be extracted from the simulated datasets. 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 Figure 4.1.

4.3.2 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, Nicola 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–$1\,h^{-1}\,\mathrm{Mpc}$ 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 (Jin et al. 2023; Cornwell et al. 2022; Kuchner et al. in prep.). If the thickness of the filaments surrounding clusters with different masses

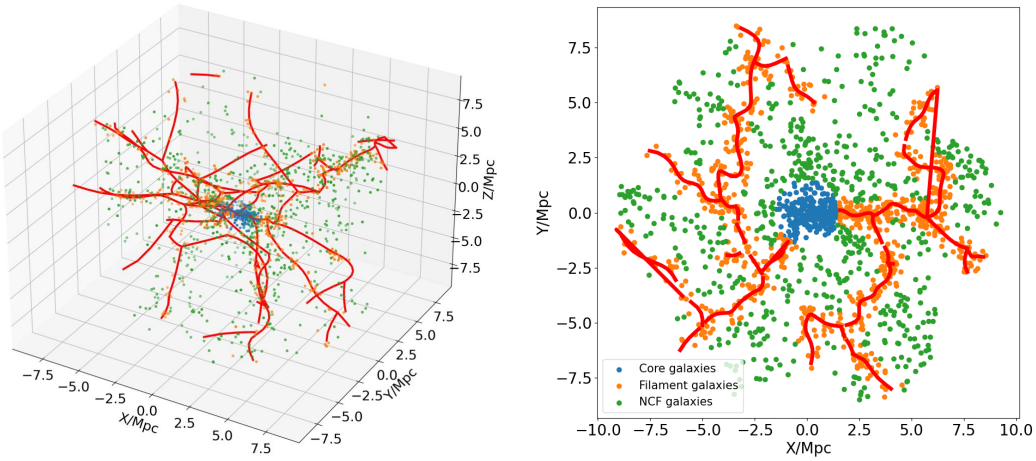


Figure 4.1: Illustration of an example cluster with the different cosmic web environments identified in 3D (directly from the simulations) and in 2D (from the mock observations based on WWFCS observational constraints). The environment assigned to a given galaxy may not be the same in 2D and 3D due to 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.

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 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 $M_{\rm cluster}^{1/3}$. Hence,

$$R_{\rm fil} = R_0 * \left(\frac{M_{\rm cluster}}{\langle M_{\rm TheThreeHundred} \rangle} \right)^{1/3}, \tag{4.1}$$

where $R_0 = 0.7\,h^{-1}\,$Mpc is the average filament radius calculated using the full TheThreeHundred cluster simulations (Kuchner et al., 2020), $\langle M_{\rm TheThreeHundred} \rangle$ is the average cluster mass in the TheThreeHundred sample ($6.5 \times 10^{14}\,M_{\odot}$), and $M_{\rm cluster}$ 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 over-simplification, but alternative recipes could be easily applied. We keep this one for simplicity throughout the chapter.

4.4 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.

Kuchner et al. (2020) showed that filamentary structures inferred from 3D galaxy positions well resemble that of the underlying dark mark matter and gas particles. As a result, we can trust that as simulations provide full 3D positional information, they allow a '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

Table 4.1: 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%

Section 4.3.1.2). 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:

1. 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 Figure 4.1 show these galaxies.

2. Filament galaxy: galaxies that lie outside the cluster core ($r > R_{200}$) and close to the filament spine ($r_{\rm spine} < R_{\rm fil}$), where $r_{\rm spine}$ is the perpendicular distance from the spine of the filament and $R_{\rm fil}$ is the filament thickness determined using Equation 4.1. These galaxies are represented by the orange points, whilst the filaments' spines are shown as red lines in the figure.

3. NCF galaxy (neither core nor filament): galaxies that lie outside of the cluster core and outside of filaments ($r > R_{200}$ and $r_{\rm spine} > R_{\rm fil}$). They are shown as green points.[2]

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 4.1. 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 $\sim 10\%$ of the galaxies, while they occupy only 1% of the volume,

[2]We note that is not an exhaustive list of cluster substructure and we focus in Chapter 5 on the detection of galaxy groups in mock observations.

and filaments contain $\sim 36\%$ of the galaxies in $\sim 6\%$ of the volume. As a comparison, over much larger regions ($\sim 500\,\mathrm{Mpc}$ scales), Cautun et al. (2014) found that filaments contain $\sim 6\%$ 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 $\sim 10\%$ of the mass but only 0.1% of the volume. Table 4.1 also shows that the density contrast between the different environments is significantly reduced when we move from 3D to 2D due to projection effects.

4.5 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 Sec. 4.2.1). This information is essential in order to interpret the observational data correctly when trying to infer how different environments affect galaxy properties and evolution.

4.5.1 Overall performance of environment allocation

In Figure 4.2, 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):

$TP =$ 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).

$TN =$ number of true-negatives = number of galaxies identified as belonging to

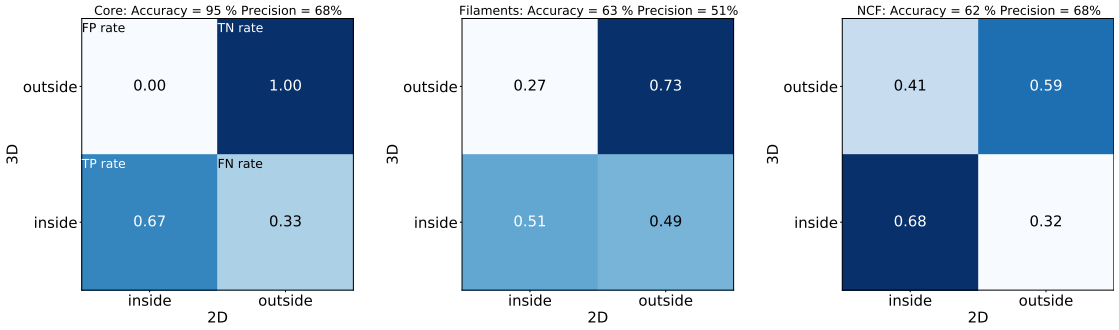


Figure 4.2: Confusion matrices comparing the environment identification for galaxies in 3D (models) and 2D (simulated mock observations). In each panel, the top left box represents the false positive rate, the bottom left box is the true positive rate, the top right box is the true negative rate, and the bottom right is the false negative rate. Left panel: galaxies identified as core galaxies. Middle panel: galaxies identified as filament galaxies. Right panel: galaxies identified as neither core nor filament galaxies (NCF). The accuracy (Equation 4.2) and precision (Equation 4.3) are displayed at the top of each panel.

environment Y ($\neq$X) in the mock observations (2D) and to environment Y in the simulations (3D).

$FP =$ 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).

$FN =$ number of false-negatives = number of galaxies identified as belonging to environment Y in the observations (2D) and to environment X in the simulations (3D).

Furthermore, we define accuracy as

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN}, \tag{4.2}$$

and precision as

$$Precision = \frac{TP}{TP + FP}. \tag{4.3}$$

In the left panel of Figure 4.2 we illustrate the success of classifying cluster core galaxies. For this environment, we get an accuracy of 95% and a precision of 68%.

For clarity, we briefly define the "rates" that are used in the panels in Figure 4.2. We use the left panel as an example:

1. FP rate: the fraction of galaxies incorrectly identified in 2D as not belonging to the cluster core. $\mathrm{FPR} = \frac{\mathrm{FP}}{(\mathrm{FP+TN})}$.
2. TN rate: the fraction of galaxies correctly identified in 2D as not belonging to the cluster core. $\mathrm{TNR} = \frac{\mathrm{TN}}{(\mathrm{FP+TN})}$.
3. TP rate: the fraction of galaxies correctly identified in 2D as belonging to the cluster core. $\mathrm{TPR} = \frac{\mathrm{TP}}{(\mathrm{TP+FN})}$.
4. FN rate: the fraction of galaxies incorrectly identified in 2D as belonging to the cluster core. $\mathrm{FNR} = \frac{\mathrm{FN}}{(\mathrm{TP+FN})}$.

This convention follows for the filament galaxies and NCF galaxies in the middle and right most panel.

The true-positive identification rate is 67%. 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%), while we are able to correctly identify galaxies not belonging to filaments in 73% of the case (true-negative rate). Conversely, we have a higher success at identifying NCF galaxies (true-positive rate of 68%) than at rejecting them (true-negative rate of 59%). 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.

4.5.2 The dependence of the environmental identification success on galaxy mass and cluster-centric 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?

4.5.2.1 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 Figure 4.1. We first divide all the galaxies in mass and 2D cluster-centric 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 $P_{\mathrm{XX}ij}$ 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, $N_{\mathrm{XX}ij}$, and the total number of galaxies allocated to environment X in 2D $N_{\mathrm{X}ij}$. In other words,

$$P_{\mathrm{XX}ij} = \frac{N_{\mathrm{XX}ij}}{N_{\mathrm{X}ij}}. \tag{4.4}$$

Conversely, we calculate the probability $P_{\mathrm{XY}ij}$ 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, $N_{\mathrm{XY}ij}$, and the total number of galaxies allocated to

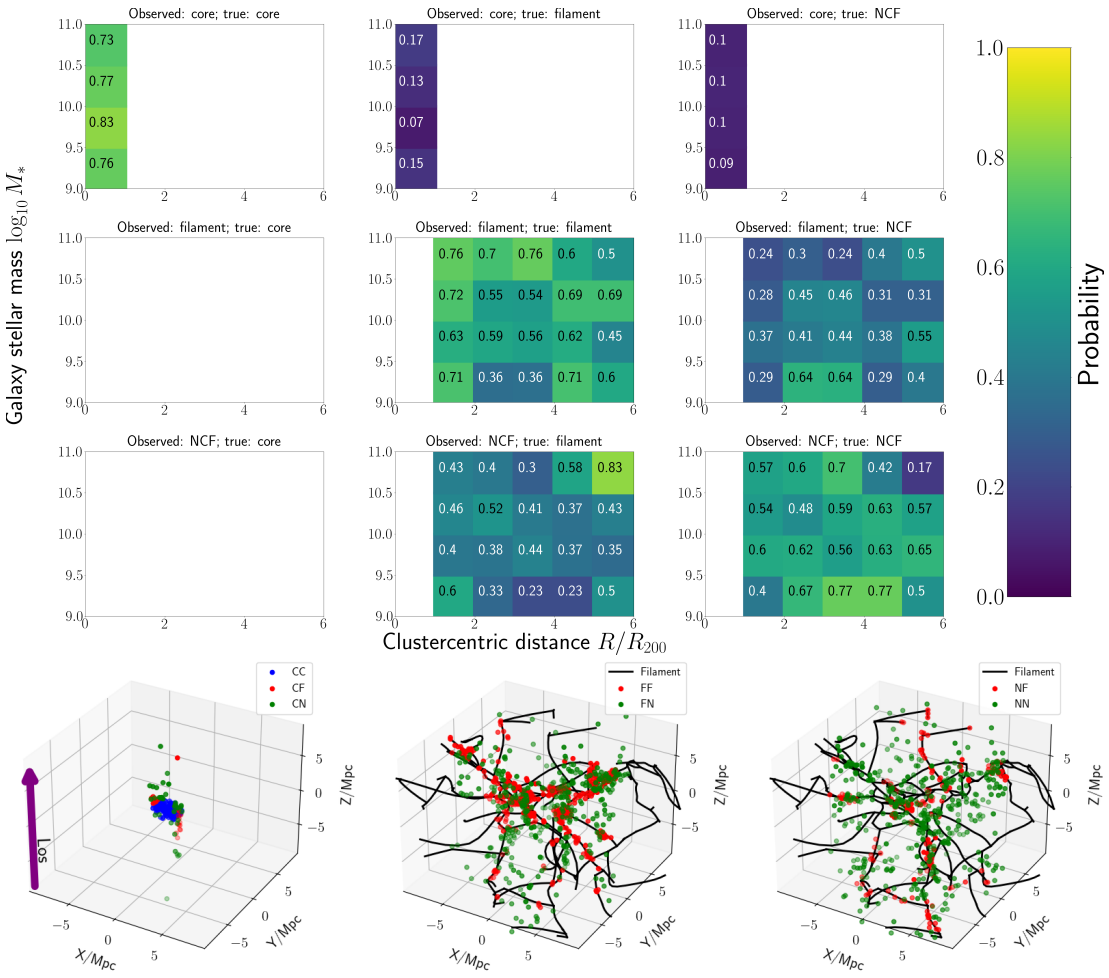


Figure 4.3: The first 9 panels (3×3 grid) display the probabilities of identifying galaxies in different environments (cluster core, cosmic filaments, or neither core nor filament, NCF)) as a function of galaxy stellar mass and cluster-centric distance in one example cluster. For every bin, we calculate the probability of correct identification $P_{\mathrm{XX}ij}$ (panels in the top-left to bottom-right diagonal) as well as incorrect identifications $P_{\mathrm{XX}ij}$ (other panels) using Equations 4.4 and 4.5. For illustration we show this for one model cluster, the same one shown in Figure 4.1. In the bottom row we show the spatial distribution of galaxies in the different environments for the same model cluster. The left 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, misidentified filament galaxies in red, and misidentified NCF galaxies 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 misidentified NCF galaxies in green. The filament network is shown as black lines. Finally, the right 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 misidentified filament galaxies in red. The line-of-sight of the cluster is indicated by the purple arrow, parallel to the z axis.

environment X in 2D $N_{\mathrm{X}ij}$. Hence,

$$P_{\mathrm{XY}ij} = \frac{N_{\mathrm{XY}ij}}{N_{\mathrm{X}ij}}. \tag{4.5}$$

We calculate this probability in four galaxy stellar mass bins and five cluster-centric distance bins, covering the ranges $10^9 M_\odot < M_* < 10^{11} M_\odot$ 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 Figure 4.3.

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, centre and right panels). For example, if we look in the $0 < r < 1R_{200}$ and $10^9 M_\odot < M_* < 10^{9.5} M_\odot$ 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 panel of Figure 4.3, where 'true' core galaxies correctly identified in 2D as belonging to the core are plotted as blue dots, those belonging to filaments in red, and those belonging to 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 Figure 4.3 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, centre, and right panels respectively). The left 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 3). The middle panel shows the probabilities of correctly classifying filament galaxies, while the right 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 4.5.2.2). The middle panel of the bottom row in Figure 4.3 shows the

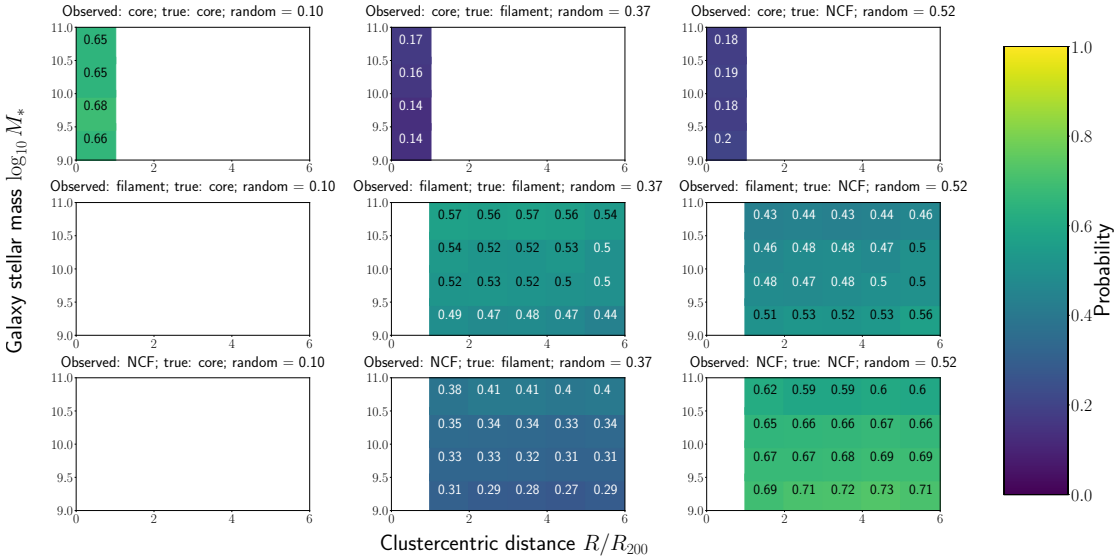


Figure 4.4: The same as the top 3×3 panels of Figure 4.3 but averaged over all clusters. The probability of randomly classifying a galaxy correctly in the title of each plot (see text for details). The information in this Figure is also included in Table B.1.

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

In a similar way, the third row of Figure 4.3 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 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 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 galaxies belonging to filaments in 2D in red. The 3D filament network is also shown as black lines.

4.5.2.2 Evaluating the entire simulated cluster sample

We repeat the process described in Section 4.5.2.1 for the whole sample of 160 simulated clusters presented in Chapter 3, see Section 4.2.2, 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 4.5.3. The average probabilities for each mass and radial bin are displayed in Figure 4.4 using the same format as Figure 4.3, with the corresponding standard deviations shown in Figure B.1.

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 re-calculating the probabilities for all clusters individually, and then averaging them. In doing so, we are using the 3D population statistics from Table 4.1 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 Figure 4.4 for easy reference. Reassuringly, the random probabilities thus calculated are very close to the true fractions shown in Table 4.1, 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 Figure 4.4 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 middle panel of the second row shows that the probability of correctly iden-

tifying filament galaxies varies from ~ 0.44 to ~ 0.57. These probabilities are always better than the random chance (~ 0.37), 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 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., 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 $\sim 17\%$ 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 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 $\sim 4\%$). 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 Figure 4.4 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.

4.5.3 The dependence of the environmental identification success on cluster mass

In Appendix B.1 (see Figure B.1) we show that there are significant cluster-to-cluster variations in the probabilities discussed above. This suggests 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 $\sim 7 \times 10^{13} M_{\odot}$ to over $\sim 10^{15} M_{\odot}$). 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 (Figure 4.5), and therefore, for the purpose of parameterising 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: $7.3 \times 10^{13} M_{\odot} < M_{\rm cluster} < 2.5 \times 10^{14} M_{\odot}$, $2.5 \times 10^{14} M_{\odot} < M_{\rm cluster} < 5.0 \times 10^{14} M_{\odot}$, and $5.0 \times 10^{14} M_{\odot} < M_{\rm cluster} < 1.2 \times 10^{15} M_{\odot}$. 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 least massive one. This is computed for every mass and radial distance bin, and shown in Figure 4.6.

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 (Figure 4.5), and therefore the fraction of the volume occupied by NCF galaxies is larger, mak-

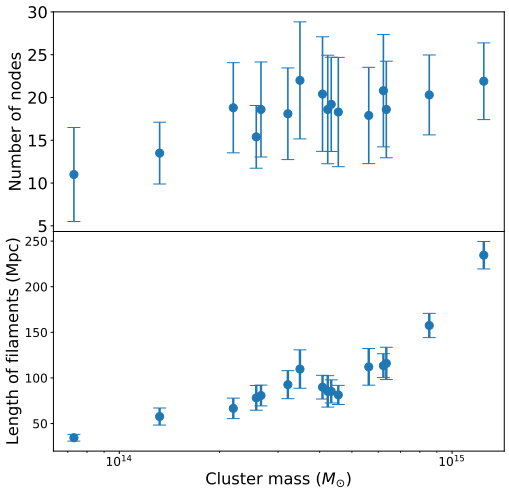


Figure 4.5: Correlation between filament network properties and the mass of the central cluster for the 160 simulated clusters mass-matched to the WWFCS sample based on `TheThreeHundred` simulations. The top panel shows the number of nodes in the network as a function of cluster mass. The lower panel 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.

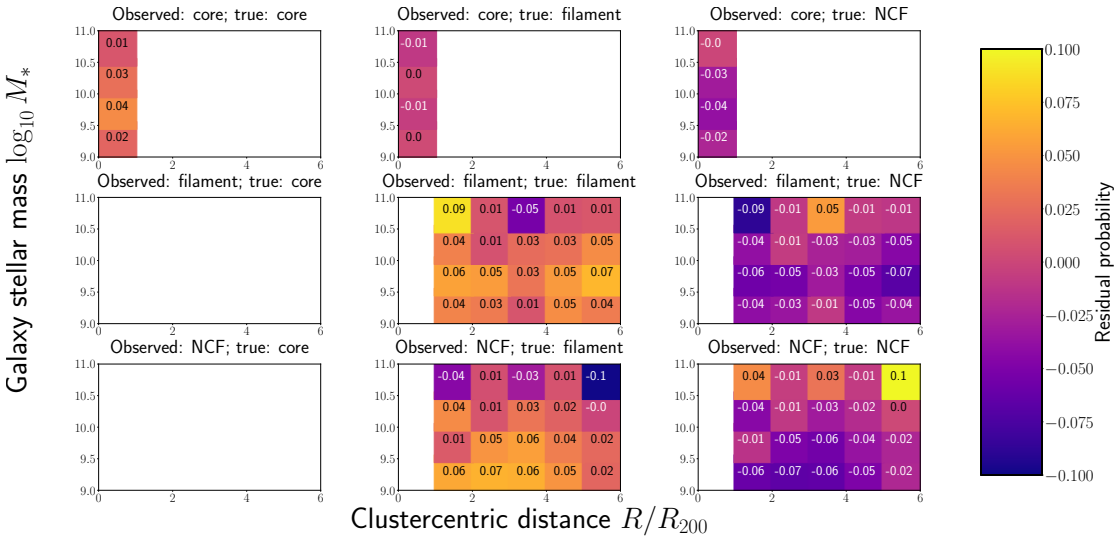


Figure 4.6: Residual probabilities of the environmental classification showing the dependency on cluster mass. These are calculated as the difference between the average probabilities for the most massive third of the clusters ($M_{\rm cluster} > 5 \times 10^{14} M_\odot$) and the least massive third ($M_{\rm cluster} < 2.5 \times 10^{14} M_\odot$). A positive residual indicates that a given probability is higher for the more massive clusters.

ing 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 Figure 4.6 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[3] (Figure 4.5), 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 Figure 4.6). 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 B.1 presents in numerical form the probabilities calculates for the complete cluster sample, as shown in Figure 4.4. Similarly, Tables B.2, B.3, and B.4 contain the probabilities for high-, intermediate-, and low-mass clusters.

[3]As a consequence of the correlation between cluster mass and total length of the filament network shown in Figure 4.5 we also find a clear positive correlation between the cluster mass and the fraction of volume occupied by filaments inside $r < 5R_{200}$.

4.6 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 3 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 chapter, 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.

1. Filaments occupy only $\sim 6\%$ of the volume enclosed in a sphere with radius $5R_{200}$ around massive galaxy clusters, but contain $\sim 38\%$ of the galaxies with masses above $10^9 M_{\odot}$. This is calculated using filament thicknesses that decrease with the mass of the main halo (see Sec. 4.3.2). In comparison, galaxies that are neither in the cluster core nor in filaments (NCF) make up $\sim 52\%$ of the galaxy population, whilst occupying 93% of the volume. The cluster core itself (defined as the sphere with radius R_{200}) contains $\sim 1\%$ of the volume and $\sim 10\%$ 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.

2. 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 cluster-centric distances, we are able to identify core, filament, and NCF galaxies with statistical accuracies (precisions) of 95% (68%), 63% (51%) and 62% (68%) respectively (see definitions in Equations 4.2 and 4.3). This indicates that, while cluster core galaxy samples can be built with a high level of complete-

ness 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. This is also explained by the method in which we calculate $D_{\rm skel}$ in Chapter 3. We note that had we calculated $D_{\rm skel}$ using the method discussed in Section 3.5.1, the lack of success in assigning galaxies to filaments would have been more expected.

3. 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 one (Figure 4.4 and Table B.1). We do that as a function of galaxy mass and cluster-centric 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 cluster-centric 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).

4. 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 B.4, B.3, and B.2), 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.

4.6.1 Caveats and future tests

There are other ways to produce the random probabilities that we use to compare our results to in Section 4.5.2.2. For example, one could only shuffle the environment tags of the filamentary and NCF galaxies given that the core galaxies are based on a pre-defined R_{200}. Furthermore, one could also randomize the environment tags within bins of galaxy mass and cluster-centric distance. We note that this choice could influence the significance of the results described in this Chapter as we would expect a larger number of high mass galaxies to appear in filamentary environments, (e.g. Sarron et al. 2019).

Chapter 5

Galaxy group detection in cluster outskirts

In the infall region of galaxy clusters, individual galaxies often exist in group sized dark matter haloes. Evidence suggests that galaxies that are being accreted onto a cluster through galaxy groups can experience environmental processing by the group itself before it reaches the virialized core of the cluster. As a result, they are regions of great interest for surveys such as the WWFCS and their characterization requires special attention. In this chapter, we investigate the identification of galaxy groups in the infall regions of galaxy clusters using the structures extractor `DisPerSE`. We investigate the co-location of galaxy groups with the critical points from `DisPerSE` and find that a high fraction of cosmic web nodes are correctly identified as galaxy groups. The fraction of matches increases with group mass and with distance from the host cluster centre. Finally, we show that when a perfect match occurs between a cosmic web node and a galaxy group, the `DisPerSE` node density δ serves as an estimate of the group's mass, albeit with significant scatter. The content of this chapter was published in Cornwell et al. (2023b).

5.1 Introduction

Some galaxies exist in galaxy cluster sized dark matter haloes. Such galaxies are subject to frequent interactions, both with neighbouring subhaloes and satellites, as well as with the intracluster medium. The impact of the environmental density on the properties of a galaxy can be clearly seen in the morphology density relation: at greater environmental densities, there is a higher fraction of early type galaxies (Dressler, 1980). Beyond the virial radius, galaxies are fed into the cluster via cosmic filaments and/or groups (Sarron et al., 2019; Martínez et al., 2015), where they experience "pre-processing" (Zabludoff & Mulchaey, 1998). Whilst the extent and location of pre-processing is still debated, recent studies at low and intermediate redshifts have shown that galaxies experience this effect before they reach first infall (Tawfeek et al., 2022; Werner et al., 2021), providing motivation for the study of the influence of filaments and groups on galaxy evolution. Next generation wide-field, multi-object spectroscopic surveys, such as the WEAVE Wide Field Cluster Survey (WWFCS; Jin et al., 2023) and the 4MOST CHileAN Cluster galaxy Evolution Survey (CHANCES; Haines et al., 2023) will directly address the need for this study. By obtaining thousands of galaxy spectra out to several virial radii around low-redshift clusters, these surveys will investigate the impact of the cosmic web around galaxy clusters and the properties of the galaxies that lie within it.

Cosmic web nodes denote areas in the large scale distribution where filaments intersect. They generally align with peaks in the density field which signal the presence of massive haloes, typically representing clusters or galaxy groups. With this in mind, Cohn (2022) used `DisPerSE`, a topological structures extractor (Sousbie, 2011; Sousbie et al., 2011), to test the matching of the location of galaxy cluster-sized haloes ($M_{200} > 10^{14} h^{-1}\ M_{\odot}$) to cosmic web nodes in the Millennium simulation (Springel et al., 2005). Using a variety of input network parameters and matching techniques, they found that 75% of galaxy clusters are matched to a `DisPerSE` node, implying that galaxy clusters represent peaks in the cosmic web. Furthermore, (Galárraga-Espinosa et al., 2024) fine-tuned their `DisPerSE` input parameters based off the matching of peaks in the Delaunay density field to massive haloes. Both of these studies were performed on cosmological box scales. However, it is unknown whether the matching of nodes to high-mass haloes ex-

tends to group-sized haloes and to scales comparable to that of the WWFCS, (regions encompassing galaxy cluster outskirts, typically out to $5R_{200}$). Furthermore, the complexity of the infall region of galaxy clusters, being the interface between the dense, non-linear cluster core and the larger-scale cosmic web, adds significant complexity to this matching.

Motivated by upcoming wide-field observations of galaxy clusters, in this chapter, we investigate the reliability of using `DisPerSE` to systematically locate galaxy groups in and around clusters by the simple process of identifying nodes in the filament network. The motivation is to encapsulate the individual components of the cosmic web (the clusters, groups and filaments) together, as one evolving field.

This Chapter is organized as follows: in Section 5.2, we describe the simulation data used in this project. In Section 5.3, we introduce our reference galaxy groups and describe the identification of cosmic web nodes. In Section 5.4 we interpret the outcome of matching cosmic web nodes to galaxy groups. In Section 5.5 we discuss the feasibility of using cosmic web node densities to interpret the mass of galaxy groups. Finally, in Section 5.6, we present our conclusions.

5.2 Data catalogues

In this chapter, we use a sub-sample of simulated galaxy clusters from TheThree-Hundred that we assembled in Chapter 3 and also used in Chapter 4. For each galaxy cluster, we take all of the haloes identified by the AHF that exceed a halo mass of $1.5 \times 10^{11} h^{-1} M_{\odot}$ which corresponds to the accumulative mass of 100 high-resolution dark matter particles. This cut lies well above the stellar mass limit of the WWFCS, (see Chapter 2 & Jin et al. 2023). We assume that in the real observations, every halo that we use in this analysis will host a galaxy.

Motivated by the WWFCS, we carry out the analysis in this chapter using both the full 3D cluster simulations and the 2D-projected cluster data: i.e., the *3D simulated clusters* correspond to the full 3D cluster region, using the x, y and z positions of the simulated haloes; and the *2D projections* correspond to the same clusters projected into 2D, using the x and y positions and omitting the z component. Kuchner et al. (2021) showed that it isn't currently feasible to

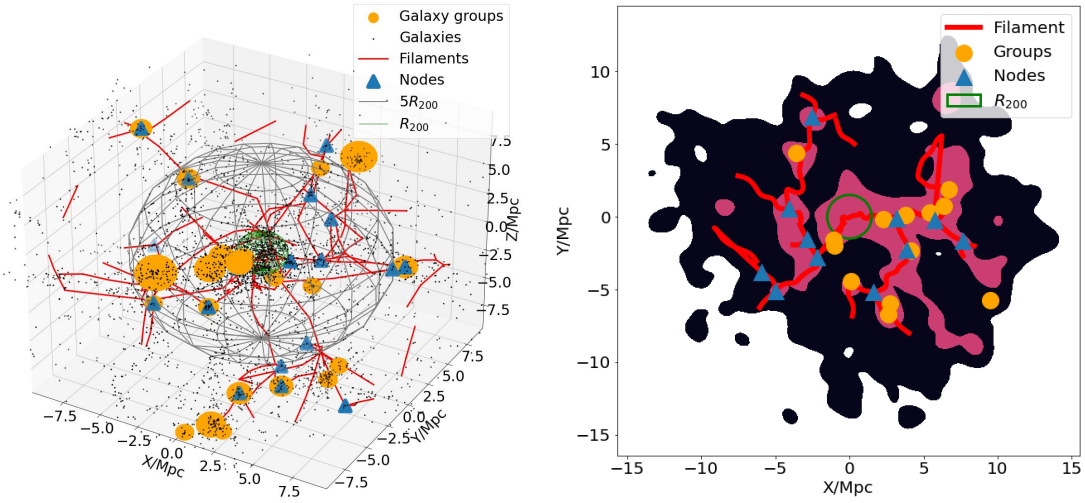


Figure 5.1: A mass-matched simulated cluster analogue of WWFCS cluster RX0058 ($M_{200} = 4.3 \times 10^{14} \mathrm{M}_{\odot}$). It corresponds to Cluster 237 from TheThree-Hundred. Left: the full 3D simulated galaxy cluster. Galaxy haloes are plotted as black dots and the 3D filament network is plotted as red lines. The larger grey sphere corresponds to $d_{3\mathrm{D}} = 5R_{200}$; the smaller orange spheres represent galaxy groups, and the central green sphere illustrates the cluster core ($d_{3\mathrm{D}} = R_{200} = 1.3$ Mpc). The nodes of this particular network outside $1.5R_{200}$ are shown as blue triangles. Right: an illustration of the 2D projection of the same galaxy cluster. A kernel density estimate (KDE) is used to represent the halo density distribution with a smoothing scale of 500 kpc. The positions of the 'true' galaxy groups and the nodes identified by `DisPerSE` are shown as orange circles and blue triangles, respectively. The filament network is shown by the red lines. The green circle corresponds to $d_{2\mathrm{D}} = R_{200}$.

reconstruct filamentary networks in 3D in surveys such as the WWFCS, hence why we project the networks in 2D for this approximation.

5.3 Identifying the large-scale structure around galaxy clusters

In the following section we discuss the identification of galaxy groups and cosmic web networks around our simulated galaxy clusters.

5.3.1 Ground truth galaxy groups

In order to address the main science question in this chapter, whether cosmic web nodes coincide with the location of galaxy groups, we firstly need to identify the 'true' galaxy groups.

Motivated by the forthcoming WWFCS observational strategy (Kuchner et al., 2022), we identify group centres in the TheThreeHundred simulations using haloes contained in the region $1.5R_{200} < d < 5.5R_{200}$, where d is the radial distance to the cluster centre. We stay well clear of the complex cluster core regions, where the peculiar velocity and the converging filament networks makes the identification of groups difficult and unreliable. This also ensures that the selected volume for all clusters is completely contained within the high resolution region and also follows the scales that the WWFCS will probe. The group haloes are then selected as objects with 1D velocity dispersion $\sigma_v > 300\ h^{-1}$ km/s (this is derived from the subhalo velocities and corresponds to a halo mass of $\sim 10^{13}\mathrm{M}_{\odot}$). In the TheThreeHundred database all galaxies within a sphere of R_{200} of the group halo are labelled as group galaxies. However, in this work, we call 'groups' the individual haloes that exceed this mass/velocity dispersion threshold, and are not concerned with group subhaloes (or group members).

In the 160 simulated galaxy clusters that we study there are 1775 galaxy groups in the 3D simulated clusters and 2430 in the 2D projections. The group catalogue used in the 2D projections is the same as the 3D group catalogue but is projected onto x and y positions. The difference in the number of groups in the 3D sample and the 2D sample stems from restricting to $1.5R_{200} < d_{\mathrm{3D}} < 5.5R_{200}$ in 3D (volume) and $1.5R_{200} < d_{\mathrm{2D}} < 5.5R_{200}$ in 2D (surface area). In the 2D sample, there are background and foreground group interlopers.

5.3.2 Cosmic web networks

We make use of the widely used structures extractor algorithm `DisPerSE` (Sousbie, 2011; Sousbie et al., 2011) to identify filaments in the simulated clusters, which is explained in detail in Section 2.3. For the purpose of this work, we reiterate that critical points identified as nodes are the maxima of the Delaunay tessellated

density field.

Previously, we have discussed the importance of mass weighting to find nodes and robust filaments using the same simulations (Kuchner et al., 2020; Cornwell et al., 2023a). Here, we apply this mass-weighting for our main analysis, but also consider the non mass-weighted networks for comparison.

In this chapter, we aim to test the matching of `DisPerSE` nodes to galaxy groups close to massive galaxy clusters. We run `DisPerSE` on the haloes that exceed the mass cut of $M > 1.5 \times 10^{11} h^{-1} M_{\odot}$. We firstly perform this test on our 3D filament networks that are ran on the 3D simulated clusters. Then, we aim to extend this to our 2D projections by running the filament networks on the 2D projected clusters (clusters projected in 2D), described in Section 5.2.

5.3.2.1 3D filament networks

In order to retrieve consistent, representative filament networks with `DisPerSE`, we need to decide on the input parameters. We choose to approximately match the number of nodes to the number of groups so that we can make a direct comparison between the two. To evaluate the matches, we define completeness and purity in the following way:

$$\text{Purity} = \frac{\text{Number of nodes matched to groups}}{\text{Number of nodes}}, \tag{5.1}$$

$$\text{Completeness} = \frac{\text{Number of groups matched to nodes}}{\text{Number of groups}}, \tag{5.2}$$

which can be computed for the 3D reference simulations and the 2D projections. In our context, the purity we calculate is used to answer: "when we find a node, how often is it truly a group?" For comparison, the completeness can be thought of as: "of all the groups that exist, how many can we find just by identifying nodes?" For example, if we use a low persistence, it is possible we will have many more nodes than groups. This will result in a high completeness but a low purity as there will be more nodes that have the potential to match to a group. By approximating the number of groups to the number of nodes, we avoid making the choice of maximizing the purity or the completeness.

For the 3D simulated mass-weighted networks, we use a persistence of 3.3σ. We use a smoothing of 5, following the work in Chapter 4, and note that smoothing does not significantly alter the positions of the critical points. To avoid a bias in our statistics, (for example, matching multiple nodes to a group), we further clean the filament networks by omitting any cosmic web node that is within 0.5 Mpc of another node, which is the case for approximately 15 nodes for the whole cluster sample (i.e., less than 1%). In this process we keep the node with the highest density field value (δ), which is the density contrast computed in the Delaunay tessellation (see below). Summed over all 160 clusters, there are 1818 `DisPerSE` nodes at clustercentric distances $1.5R_{200} < d_{\rm 3D} < 5.5R_{200}$, compared to 1775 groups.

5.3.2.2 2D filament networks

In order to make a direct link to observations (see Kuchner et al. 2021 and Chapter 3 for details), we produce filament networks using the 2D projections of the simulated galaxy clusters. We run `DisPerSE` on the x and y positions of the haloes and apply mass-weighting to construct the 2D projected filament networks. Here, we use a persistence of 2.7σ and a smoothing of 5. After cleaning the networks in the same manner as described above, we produce 2327 nodes, compared to 2430 groups in the range $1.5R_{200} < d_{\rm clus,2D} < 5.5R_{200}$.

We note that when observing real cluster regions we don't know *a priori* the true number of groups, which we have used to set the persistence value. As we have done here, one can use simulations to estimate the expected number of groups for clusters of a given mass, and use that to set the persistence value to derive the observed filament network.

An example simulated galaxy cluster is shown in Figure 5.1. In the left panel, we see the full 3D cluster with the filament network (derived using mass-weighting) overlaid in blue. Galaxy groups are illustrated as orange spheres where their radius corresponds to R_{200} of the group itself. Cosmic web nodes are shown as blue triangles. We show the corresponding 2D projection of the same cluster in the right panel. Here, a kernel density estimate is used to display the cluster density field, and the filament networks are plotted in red. As before, we show the

positions of the galaxy groups and cosmic web nodes as yellow circles and blue triangles.

5.3.3 Node and group number densities

To test whether `DisPerSE` nodes match with galaxy groups around clusters, we first compare their number densities as a function of radius, both in 3D and in projected 2D. To do this, we calculate the volume and surface number density of groups and nodes in a range of concentric shells for each cluster, where the volume and surface number densities are calculated in units of $(r_{\rm 3D}/R_{200})^{-3}$ and $(r_{\rm 2D}/R_{200})^{-2}$ respectively.

The number density of groups and nodes follows a monotonic decline with clustercentric distance both in 3D and 2D (Figure 5.2). Close to clusters, the number density of nodes and groups are different, both in 3D and 2D: in the innermost radial bins we identify significantly fewer nodes per unit volume (area) than groups. We therefore expect that this mismatch may affect the completeness of the matching of nodes to groups near the cluster cores. This discrepancy is especially relevant in light of our initial decision to approximately align the total numbers of nodes and groups. We interpret this as due to the cluster core dominating the local density field and thereby diminishing the likelihood of persistence pairs forming close to the cluster core. Beyond $2.5R_{200}$, the number density of nodes and groups begin to converge and agree within each other's standard error. We discuss the implications of these results in more detail in the following section.

5.4 Results

Motivated by surveys such as the WWFCS, we investigate whether cosmic web nodes as detected with `DisPerSE` match to galaxy groups in the outskirts of galaxy clusters. We carry out this analysis with haloes in cluster simulations in 3D and in projected 2D.

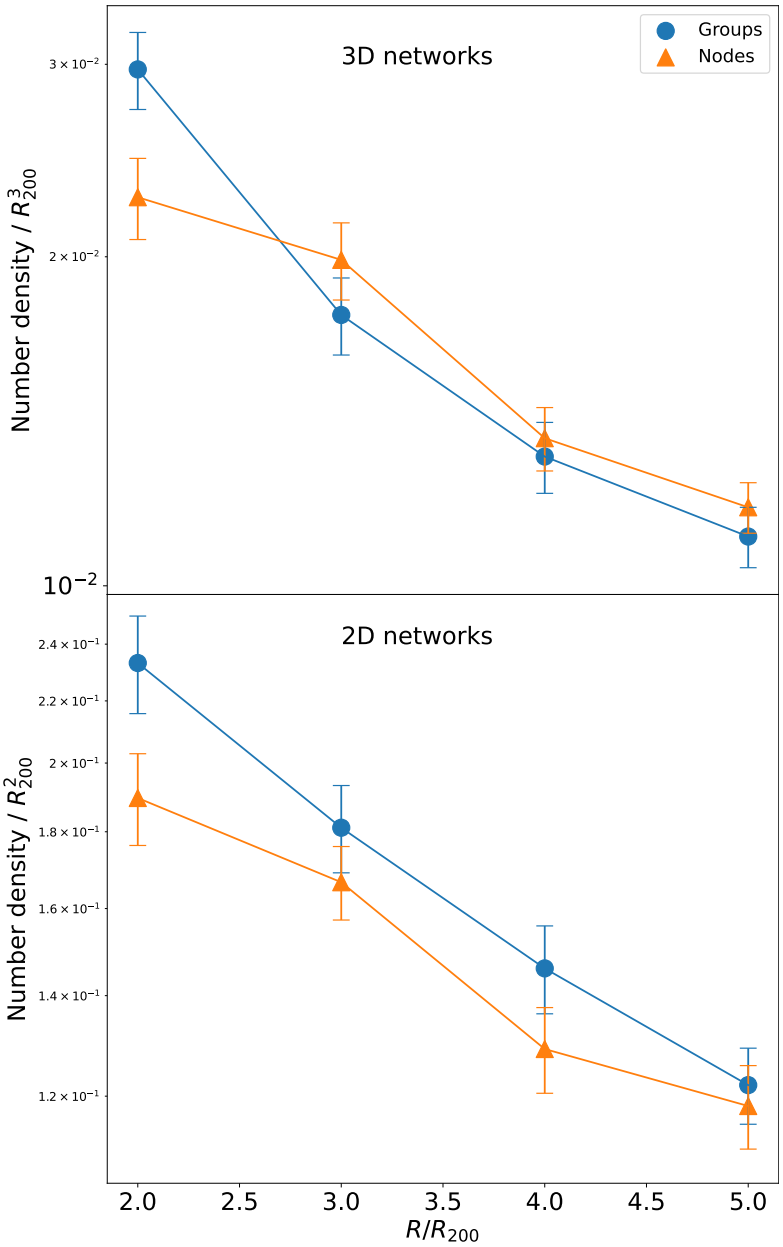


Figure 5.2: The radial number density of galaxy groups and cosmic web nodes decrease as a function of cluster distance. The top panel shows results from mass-weighted `DisPerSE` 3D filament networks, the bottom from projected 2D networks. The points show the mean number densities and the error bars are the corresponding standard errors. We find the largest discrepancy between nodes and groups at small clustercentric distances.

5.4.1 Matching groups to nodes in 3D simulations

To test the coincidence between nodes and groups, we compute the nearest neighbour from every node to every group in each individual cluster. For there to be a successful match, we require that the node be within a radial distance of R_{200} of the corresponding group centre. Where there are matches between multiple nodes and groups, we take the node with the highest density as calculated by `DisPerSE`.

Figure 5.3 quantifies the success in matching cosmic web nodes to galaxy groups in the mass-weighted case. The top left panel of Figure 5.3 shows the relationship between distance of a node to a galaxy group and the group mass, represented by a color map. The green dot-dashed vertical line represents the upper limit of what we label a 'perfect match'. Anything considered a 'perfect match' is where the distance between a node and a group is essentially zero, i.e., below the simulation resolution limit. The dashed red line illustrates the boundary of what we consider to be a close match to a group: this corresponds to R_{200} of the respective group that we are analyzing. There are two main peaks in the distribution, as replicated in the lower left panel, with the dominant peak in the 'perfect match' range and the secondary peak corresponding to a slightly lower instance of non-matches. For the 3D simulated clusters, out of the 1818 nodes, 1011 are associated with galaxy groups (56%). The lack of node–group pairs in Figure 5.3 with distances in the $10^{-4} \lesssim R/R_{200} \lesssim 10^{-1}$ range is due to mass-weighting and the fact that nodes are always located at the centre of a halo: if a less massive halo resides very close to a group-mass halo ($R/R_{200} \lesssim 10^{-1}$) the node would 'latch' to the group-mass halo itself, and not the lower-mass halo.

For illustration, in Figure 5.4 we display the filament network from one model galaxy cluster in our sample. In the top left panel we show the network in black, with the nodes as blue triangles and groups as orange discs. We also show the other `DisPerSE` critical points, saddle points (local density minima) and bifurcations (where two or more filaments intersect without a maximum being present). The other panels then zoom in on three different regions that exemplify a 'no match' between a node and a group, a 'close match' and an instance of a 'perfect match', enclosed by a red, yellow and green mesh sphere. In our example in the top right panel, `DisPerSE` does not place a node where a group is. Instead, a node

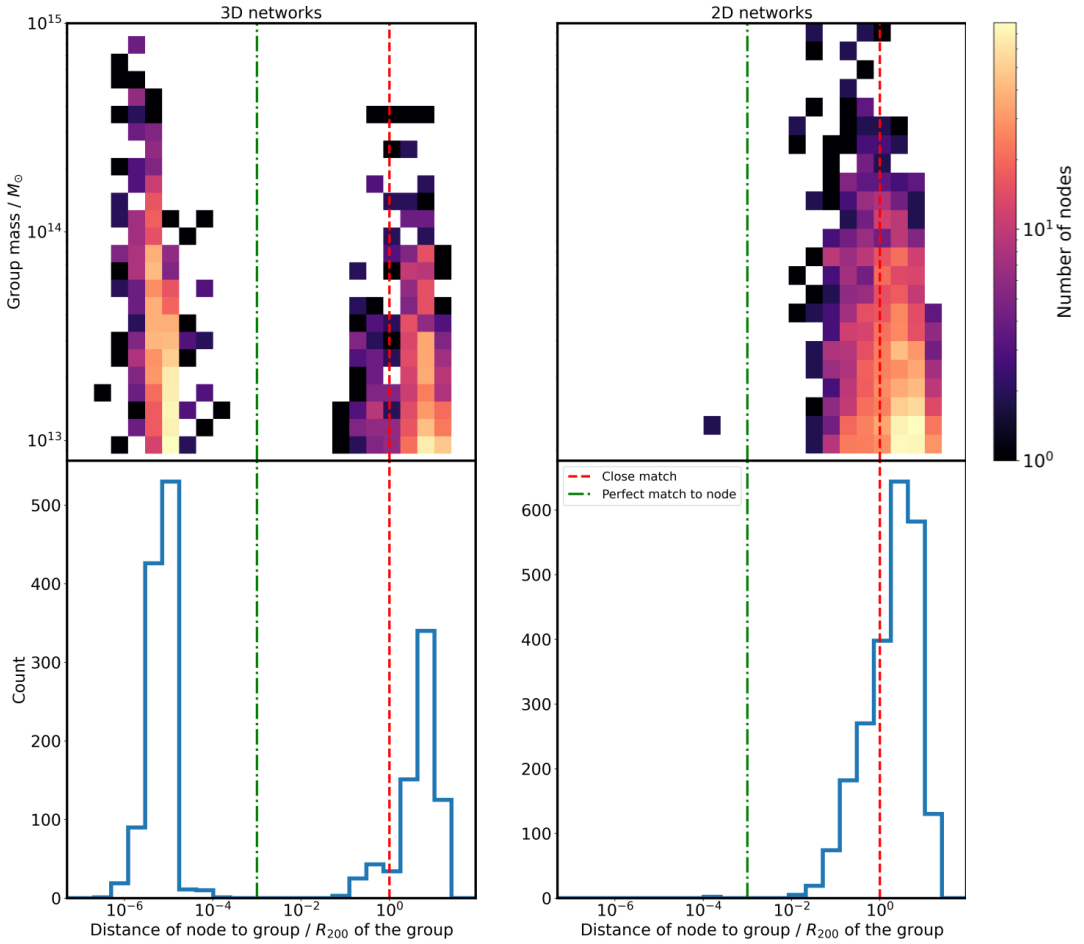


Figure 5.3: An illustration of the level of success between the matching of cosmic web nodes and galaxy groups in the 3D reference simulation (left) and the 2D projections (right). Top two panels: a colour map showing the normalized distance (in units of R_{200}) of each node to its nearest group against the mass of the group in the 3D simulations (left) and the 2D projections (right). The colour corresponds to the density of points. The dashed red line represents our criteria for a close match and the dot-dashed green line corresponds to a perfect match. The bottom two panels show the corresponding 1D histogram of the distance of each node to a group. In 3D, most cosmic web nodes successfully match to a galaxy group. However, the link is significantly weakened in 2D.

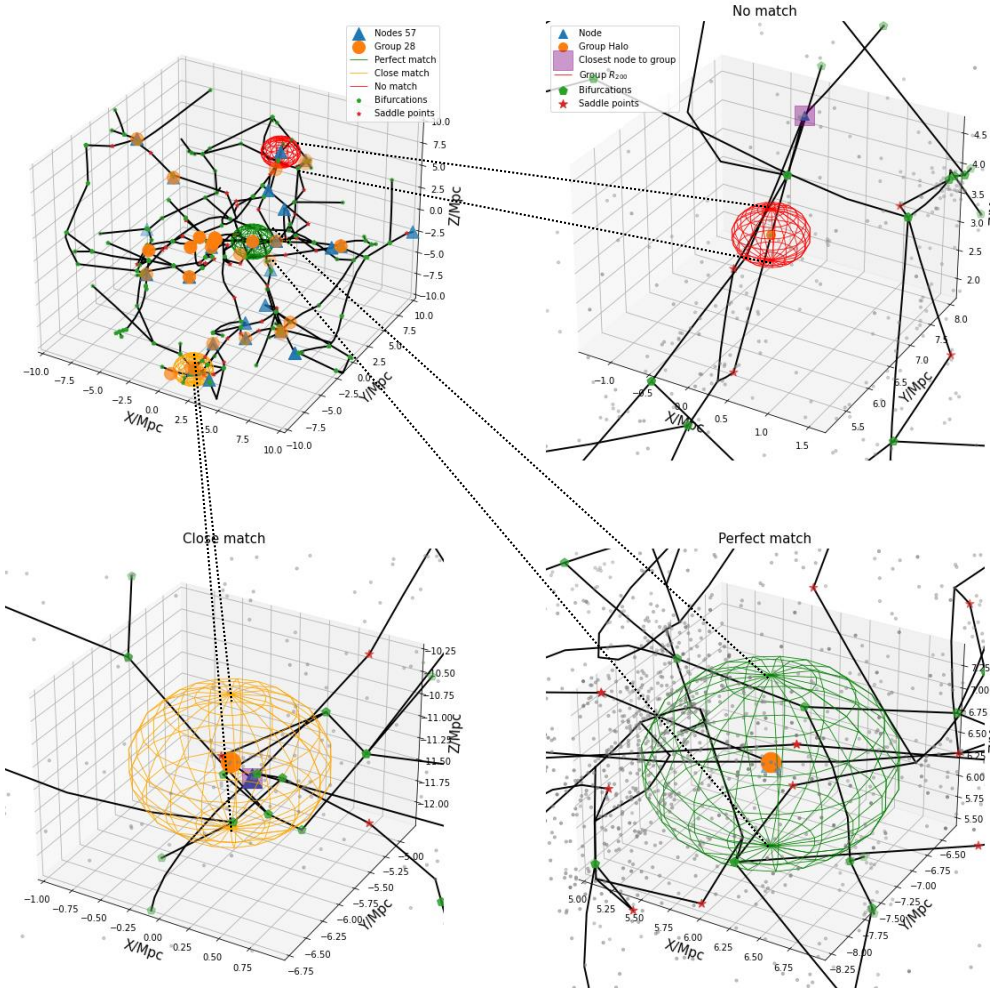


Figure 5.4: An example galaxy cluster ($M_{200} = 4.3 \times 10^{14} M_{\odot}$) with the `DisPerSE` critical points and group haloes overlaid. The top left panel shows the full smoothed filament network in black. The cosmic web nodes are shown as blue triangles, the bifurcation critical points are displayed as green pentagons, and the saddle points are illustrated as red stars. We have also displayed the galaxy groups as orange circles for reference. As explained in the text and demonstrated in Figure 5.3, we calculate the distance between each node and each galaxy group within the cluster and show an example of a non-match, a close match and a perfect match as red, yellow, and green mesh spheres respectively. The top right panel illustrates a zoom in on a 'non-match' between a node and a galaxy group, where we have also plotted the haloes from the simulation. The bottom left panel is a zoom in on an example of a 'close match' and the bottom right panel shows an example of a 'perfect match'. In the top right and lower left panels we have plotted the closest node to a group, showing where a cosmic web node has not latched on to the closest group. The radius of the mesh spheres in the upper right and lower two plots correspond to R_{200} of the group halo.

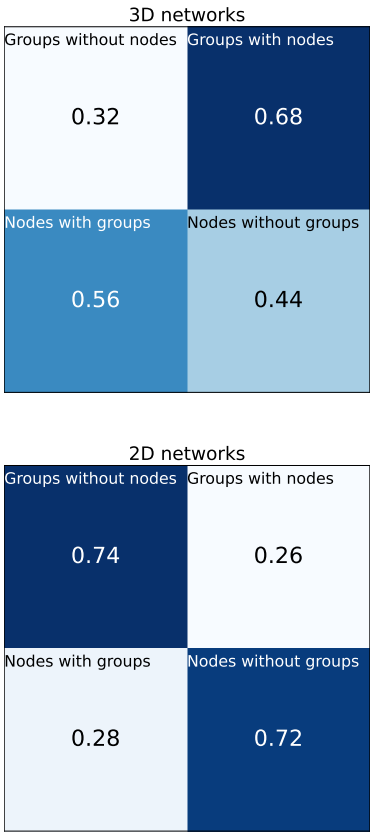


Figure 5.5: A confusion matrix illustrating the relative success of the matching of cosmic web nodes to galaxy groups in the 3D simulations (top panel) and the 2D projections (lower panel). The bottom row of each matrix is calculated by the number of nodes with/without groups divided by the total number of nodes. The top row is calculated as the number of groups with/without nodes divided by the total number of groups. We only consider groups and nodes in the region $1.5R_{200} < r < 5.5R_{200}$.

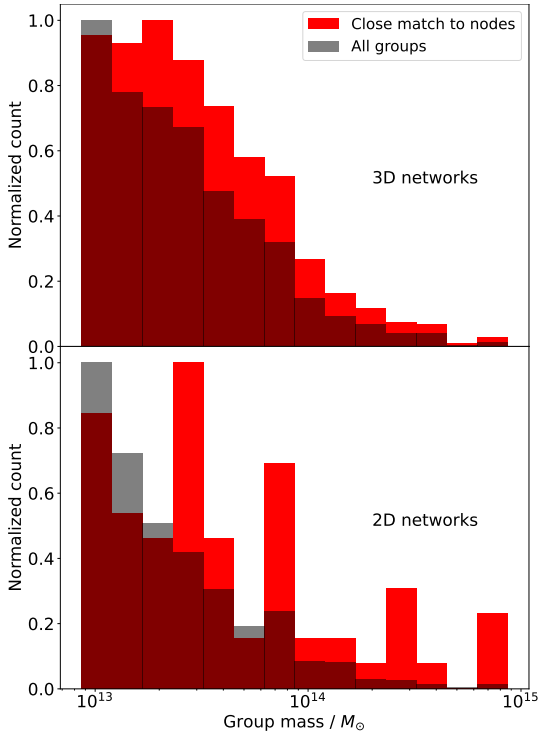


Figure 5.6: Mass distribution of all galaxy groups and those groups considered successful matches to nodes. We present the results for the 3D simulated clusters in the top panel and the results for the 2D projections in the lower panel. We display the normalized histograms of the group masses in black and the histogram for all galaxy groups that are close matches to nodes in red. The mass distributions appear significantly different, such that in the 3D reference simulations, groups that are close matches to nodes tend to be more massive than the general group sample.

is identified at a distance of 1.75Mpc. In the bottom left panel, the node has been associated with a halo that is a subhalo of a galaxy group, but not the main group halo, and it is therefore a successful match but not a perfect match. This corresponds to the data points between the dash-dotted green line and the dashed red line in Figure 5.3. This demonstrates that while mass-weighting helps, it does not always result in a direct match from the galaxy group halo to a node. In the bottom right panel, the cosmic web node has latched on to a group-sized halo.

We quantify the success of matching nodes and groups using a confusion matrix in Figure 5.5. Summing over all of the clusters, we calculate a purity of 56% and a completeness of 68%. Whilst we have statistically demonstrated that there is a link between the positions of cosmic web nodes and galaxy groups, we note that there remains significant contamination, with 44% of nodes not matching to groups and 32% of groups not matching to nodes. [1]

In the top panel of Figure 5.6 we show the normalized mass distribution of the entire sample of galaxy groups as well as the mass distribution of galaxy groups that are close matches to nodes. More massive groups are more likely to match `DisPerSE` nodes. To quantify this, we perform a Kolmogorov-Smirnov test, which is a non-parametric test of the equality of two continuous, one dimensional probability distributions. We test the null hypothesis that the cumulative mass distribution of groups could be drawn from the 'close match' to node group mass distribution and chose a significance threshold of 0.05. We record a p-value of 10^{-6} which allows us to reject the null hypothesis. In other words, galaxy groups that are located at cosmic web nodes have an intrinsically different mass distribution to that of the general group population: they are typically more massive. This result agrees with Cohn (2022). They further found that matched clusters to nodes tend to occur in nodes of higher density (correlated with cluster/group mass, see below), and have a slightly less recent major merger. This is something we will explore in Chapter 6.2.

[1] We note that the purity and completeness are largely influenced by the persistence. By increasing the persistence, there will be less critical points and therefore, less nodes. In turn, this would decrease the completeness but increase the purity.

5.4.2 Matching groups to nodes in 2D projections

We expect that matching nodes to groups is more challenging in projected 2D. One of the obvious reasons is that we are losing 1/3 of the spatial information when we project the simulated cluster volumes. On the other hand, it is possible that in the 2D projections, we may produce false matches between nodes and groups. This is where a node is close to a group 'on the sky', meaning they are a match in the 2D projections, but their real line-of-sight distance is large and would result in a non-match in the 3D simulations. With this in mind, in this section we investigate the success and limitations of the matching of cosmic web nodes to galaxy groups in 2D projections.

The right panels of Figure 5.3 illustrate that the matching of cosmic web nodes to galaxy groups is significantly different compared to the 3D reference simulations. There is one main peak in the distance distribution that straddles the boundary of a 'close match' but lies preferentially in the 'no match' region. This is echoed in the bottom panel of Figure 5.5, where we present the purity and completeness. Overall, we find that of the 2327 cosmic web nodes, 662 of them match to galaxy groups (a purity of 28%). The corresponding completeness is significantly worse than the 3D case and is calculated to be 26% compared to 68%. In the bottom panel of Figure 5.6, we display the mass distributions of the galaxy group sample and the mass distributions of those groups that are close matches to nodes. We perform the same KS test, using the same null hypothesis and significance threshold and calculate a p-value of 10^{-5}. We note that the mass distribution of groups here is contaminated by the projection into 2D and may add to the spurious peaks seen in the distribution.

As expected, the matching between nodes and groups in 2D returns a lower purity and completeness than in the 3D simulations. The dimensional reduction severely impacts the success in matching. However, Figure 5.2 shows a convergence in the number density of nodes and groups at greater distances from the cluster core. With this in mind, in the next section we examine the matching of cosmic web nodes to galaxy groups as a function of clustercentric distance.

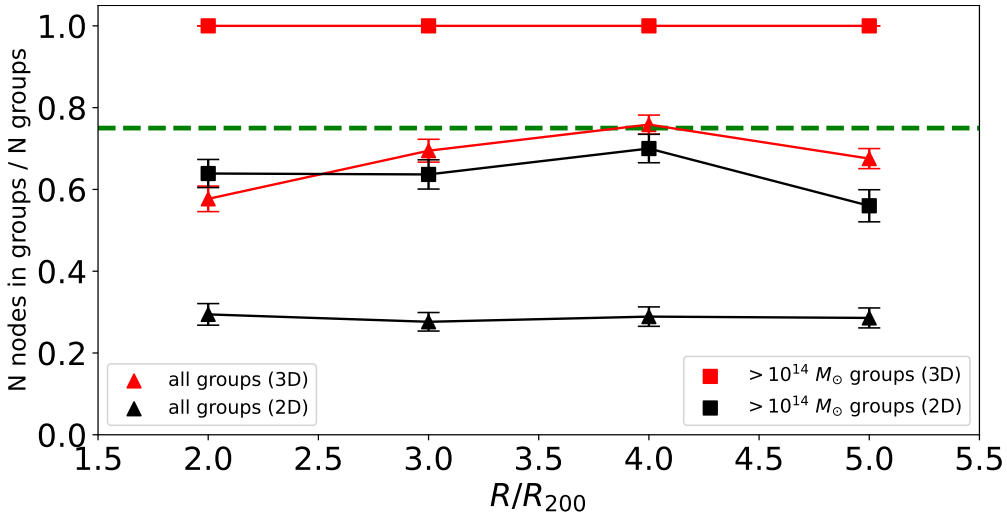


Figure 5.7: Fraction of groups in nodes (completeness) as a function of clustercentric radius, summed over our entire 3D simulated cluster sample and the 2D projected sample. We evaluate this fraction at four intervals from $1.5R_{200}$ to $5.5R_{200}$. The red triangles display the completeness in matching the entire group sample in 3D and the black triangles illustrate the completeness in 2D. We show the case where we limit to the most massive galaxy groups as red squares in 3D and in black squares in 2D. Error bars indicate the standard error. The green dashed line shows the average fraction obtained over much larger scales (Cohn, 2022) for comparison. In 3D, cosmic web nodes match well to very massive galaxy groups, with a success rate close to 100%.

5.4.3 Radial dependence on matching nodes to groups

5.4.3.1 3D filament networks

We start by testing the radial matching of cosmic web nodes to galaxy groups using the same radial bins discussed in 5.3.2.1 and present the findings in Figure 5.7. Here, the y axis is a measure of the fraction of cosmic web nodes that reside in galaxy groups divided by the total number of groups in that radial bin – the completeness. This is calculated for each cluster and the mean is represented by the red triangles with the error bars representing the standard error.

Generally, for the entire group catalogue, the success rate improves as we increase the clustercentric distance. This is to be expected – close to the cluster core the main halo dominates the density field and therefore prevents the formation of critical points that exceed our persistence threshold. As the gravitational influence

of the cluster decreases, more persistence pairs can form and are therefore more likely to align with the high mass groups. Encouragingly, as the clustercentric distance increases, the success rate approaches that of Cohn (2022), although we must bear in mind that the mass range of the haloes we use (10^{11} $M_\odot$ < $M_{\rm halo} < 10^{15}$ $M_\odot$) is much larger than the one used by this author ($M > 10^{14}$ $M_\odot$); furthermore, we are probing very different distance scales (tens of Mpcs compared to hundreds of Mpcs). Nevertheless, our relatively high success at matching nodes and groups in the complex vicinity of clusters (at least in 3D) seems promising. We note that the mass-weighting scheme we use in `DisPerSE` plays a very important role (cf. Appendix A) and is largely responsible for the success in matching nodes to galaxy groups.

In Figure 5.7 we also display the success rate for massive galaxy groups, using the same mass threshold as the Cohn (2022) work ($M > 10^{14}$ $M_\odot$). Limiting the sample to these high mass groups, we find that completeness jumps to 100%. This implies that the derivation of `DisPerSE` nodes can be used in the detection of nearly all massive galaxy groups in close proximity to clusters when one applies mass-weighting.

5.4.3.2 2D filament networks

We repeat the analysis for the 2D projected networks. Contrary to the results of Section 5.4.3.1, we find that there is no improvement in the success rate with increasing radius, resulting in a flat rate of approximately 26%. However, when we consider only the higher mass groups, we find a stark improvement in the matching of nodes to galaxy groups: 63% of $M > 10^{14}$ $M_\odot$ groups match to a cosmic web node. Interestingly, there appears to be little dependence on the clustercentric distance.

We conclude that the effect of the large contamination rates in the purity and completeness of the node-matched group sample prevent us from using `Disperse` to identify a robust sample of galaxy groups. However, we have shown that our approach is much more successful when considering only the most massive groups. Therefore, we expect to be able to identify $\sim$ 63% of all galaxy groups with $M > 10^{14} M_\odot$ using nodes identified by `DisPerSE`. Furthermore, for all group

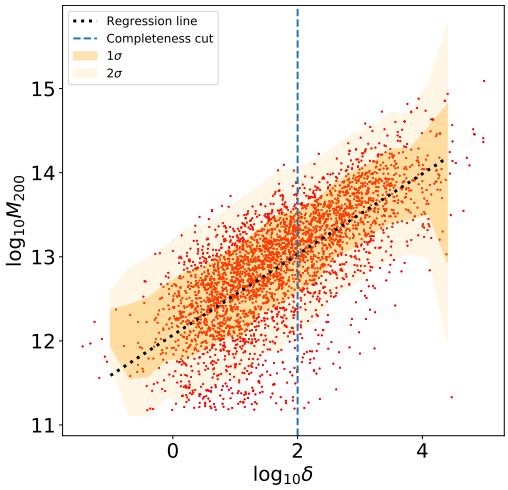


Figure 5.8: Correlation between the mass of a galaxy group and the density of the node that it matches to in the 3D reference clusters (red dots; see text for details). The dotted black line shows the regression line obtained by fitting only the data in the complete sample, as discussed in the text. The sample is complete for $\log_{10}\delta > 2$ (indicated by the blue dashed line). The 1σ and 2σ scatter are overlaid using two different shades of yellow. The node density of the cosmic web can be used to estimate the mass of a galaxy group, albeit with large scatter.

masses, `DisPerSE` can be applied to observations to locate potential galaxy groups using cosmic web nodes, and then the resulting sample can be verified and cleaned using alternative, perhaps more ad-hoc and less systematic methods.

5.5 Group mass estimation from cosmic web node density

We have shown that it is possible to locate a significant fraction of galaxy groups using cosmic web nodes in 3D. We now examine whether we can use `DisPerSE` to estimate the masses of these groups. In what follows, we demonstrate that the node density of the cosmic web provides information on the masses of galaxy groups that are associated with them.

As mentioned in Section 5.3.2.1, `DisPerSE` outputs a list of density values that are calculated during the Delaunay tessellation for each identified critical point. We have demonstrated that, in 3D, there is a tendency for cosmic web nodes to match

to galaxy groups and we therefore further investigate the possibility of using the density of the nearest node to a galaxy group in order to estimate its mass. We do this by taking the galaxy groups that are a 'perfect match' to a node and record its corresponding node density. In Figure 5.8 we show that there is a strong positive correlation between these two parameters, albeit with considerable scatter.

We then fit a least-squares regression line to the sample of nodes for which a complete mass-selected sample of corresponding groups can be identified in the simulations. Visual inspection of Figure 5.8 indicates that groups with masses that correspond to node densities of $\log_{10}\delta > 2$, indicated by the blue dashed line in the figure, constitute such a complete sample. We only fit a regression line to that complete sample to avoid any Malmquist-like biases. The figure shows that an extrapolation of this line to lower group masses is a fair representation of the trend at all masses. The equation of the regression line is

$$\log_{M_{200}} = 0.48 \log_{10}\delta + 12.07. \tag{5.3}$$

Using this regression line it is possible to obtain a rough estimate for the mass of the groups selected as `DisPerSE`-identified nodes, but the scatter is large, roughly a factor of ~ 3.

5.5.1 High mass groups and node density matching

We have shown that there is a strong positive correlation between the node density and mass of its closest matched halo. It is therefore possible to select the highest mass groups ($M > 10^{14}$ $\mathrm{M}_{\odot}$) by finding a suitable cut in the node density. In this section, we test the matching of high mass groups to high density nodes in order to uncover whether this method is a robust strategy for detecting massive galaxy groups.

Using equation 3 we find that a group mass of 10^{14} $\mathrm{M}_{\odot}$ corresponds to $\delta \sim 10000$ in the node density. By considering only the highest mass groups and the nodes above this density threshold, we repeat the analysis in Section 5.4.1 by quantifying the matching of high mass groups to high density nodes. We note that we exclude clusters from this analysis that do not match to any high mass galaxy groups.

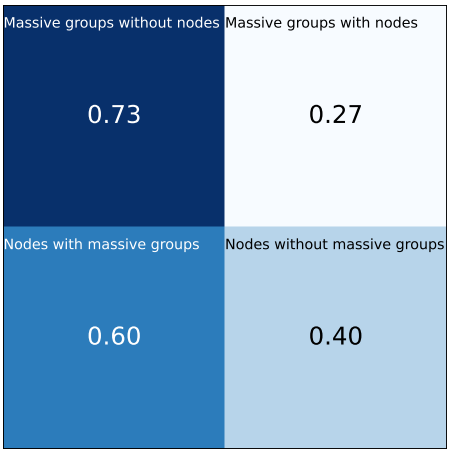


Figure 5.9: Same as Figure 5.5 for the 3D reference simulations but for high mass groups ($M > 10^{14} M_{\odot}$) selected as high density nodes, as described in Section 5.5.

Figure 5.9 illustrates the results from this analysis. We find that the fraction of high-density nodes in high mass galaxy groups (purity) is 60%, very similar to that of the entire group sample. However, the fraction of high mass groups matching high density nodes (completeness) decreases to 27%. We attribute this to there being a greater number of groups (89) than nodes (42) above the mass and density thresholds. When we limit the node density, we exclude some cases where nodes match with galaxy groups but lie below the density threshold, thereby negatively affecting the matching of massive groups to high-density nodes. We conclude that, although there is a strong correlation between the group mass and cosmic web node density, restricting the node density in this way does not significantly improve purity or completeness.

5.6 Conclusions

Galaxies experience different physical processes in different cosmic web environments. Next generation wide-field spectroscopic surveys will, for the first time, be able to accurately map the distribution of galaxies to cosmic web features around statistical samples of galaxy clusters, where pre-processing is present. In this chapter, we present and evaluate a novel approach for identifying galaxy groups (haloes with $\sigma_v > 300 h^{-1}$ km/s) near massive galaxy clusters utilizing the critical points identified as network nodes using the `DisPerSE` software, (which we run

on haloes with more than 100 dark matter particles). We summarize our main findings below.

1. We have tested the matching of cosmic web nodes, (derived from mass-weighted filament networks from `DisPerSE`), to galaxy groups in and around massive galaxy clusters in 3D and projected 2D cluster simulations. We find a purity of $\sim 56\%$ and a completeness of $\sim 68\%$ in 3D and a purity of $\sim 28\%$ and a completeness of 26% in 2D.

2. We find that the galaxy groups that closely match with cosmic web nodes tend to be the more massive ones.

3. In the 3D reference simulations, we find a slight improvement in the fractional number of nodes within galaxy groups as we move further away from the cluster core. This suggests that the cluster core's complexity hinders the accurate matching of density field maxima to galaxy groups. Conversely, in the cluster outskirts the success rate increases due to the dominant influence of the large scale cosmic web rather than the cluster core. Within the range of 3–5R_{200}, the number of nodes in groups reaches a maximum of approximately $\sim 75\%$, matching the results obtained by Cohn (2022) over much larger spatial scales. In contrast, in the 2D projections, we do not see any radial trend.

4. Limiting our analysis to only the most massive galaxy groups ($M > 10^{14} M_{\odot}$), we find that 100% of cosmic web nodes match to galaxy groups in the reference 3D simulations. We also find a stark improvement in the success rate for 2D projections, increasing from $\sim 28\%$ to $\sim 63\%$.

5. We find a strong positive correlation between the mass of the groups and the `DisPerSE`-determined density of their matching nodes. This correlation (equation 3) can be used to obtain a rough estimate of the group mass within a factor of ~ 3.

In summary, we have shown that the widely-used topological filament finder `DisPerSE` can be used as a powerful tool for identifying galaxy groups around clusters. It can be further complimented with other group finding algorithms. We have tested the strengths and weaknesses of this approach with future wide-field

surveys of galaxy clusters in mind. While we used simulated galaxy clusters in this study, its accuracy can be scrutinized with diverse group-finding methods and observational data, thus opening new avenues for the study of galaxy groups and their role in galaxy evolution.

Chapter 6

Conclusions and Future Work

Throughout this thesis, we have used the hydrodynamical simulations from The ThreeHundred, coupled with the topological structures extractor `DisPerSE`, to test and optimize our ability to extract different cosmic web environments. This work began with an investigation into our ability to reconstruct cosmic web filaments using mock observations of galaxy clusters, matched to clusters observed with the WWFCS. This allowed us to verify that despite the selection effects of the WWFCS, we can accurately map the pathways that cosmic filaments trace. Following this, we then tested our ability to associate individual galaxies to different cosmic web environments. Finally, we investigated the possibility of detecting galaxy groups in close proximity to clusters using `DisPerSE`. We found that future surveys can use this framework for detecting massive galaxy groups in the outskirts of galaxy clusters, whilst simultaneously being able to extract robust filamentary networks.

In this chapter, we summarize these results in more detail, and discuss future implications of the work in this thesis.

6.1 Summary of results

In this section, we summarise the work from Chapter 3, Chapter 4, Chapter 5.

6.1.1 Filamentary networks in galaxy cluster simulations

We begin this thesis with a study into the feasability of extracting filamentary networks around galaxy clusters. Currently, there is evidence suggesting that filamentary environments can act to suppress star formation in galaxies (e.g. Hoosain et al. (2024)). There is further observational evidence suggesting filaments can act to enhance the HI content of galaxies that reside in them (Kleiner et al., 2016). Next generation wide-field spectroscopic surveys, such as the WWFCS and 4MOST CHANCES, will aim to probe the impact of cosmic web filaments on galaxy evolution close to galaxy clusters. Therefore, understanding our ability to extract cosmic web filaments by creating mock observations is an important test.

Using TheThreeHundred simulations of galaxy clusters, we mass-matched 10 galaxy clusters to each individual cluster that will be observed in the WWFCS. After applying the projection effects and the instrumental selection effects, we took the last steps in designing fully informative mock observations from simulations. Our findings included:

1. After 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 $72.7\% \pm 1.7\%$ 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.

2. 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 $D_{\rm skel}$ in the reference and recovered networks is only $0.13 \pm 0.02\,$Mpc on average, an much smaller than the

typical filament radius of $\sim 1\,\mathrm{Mpc}$. Furthermore, only $\sim 11-13\%$ of all recovered filament segments lie at a distance larger than 1 Mpc away from their corresponding reference segment.

3. 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.

4. We discuss the possibility of recalculating D_{skel} by omitting the segments inside R_{200}. As our current D_{skel} distributions are likely biased by the higher concentration of segments inside of R_{200}, this would provide a better test of the mapping of filaments in our "mock observational" networks to the "simulated" networks in the outskirts of clusters. In doing this, our median D_{skel} values would likely be larger. We note that this may also better explain the lack of success we had in Chapter 4 in assigning galaxies to filaments.

This chapter enabled us to confirm that next generation surveys will be able to extract high-quality filamentary structures close to galaxy clusters.

6.1.2 Allocating galaxies to cosmic web environments

After the initial work in designing, optimizing and testing the observational strategy in Chapter 3, we test our ability in allocating individual galaxies to different cosmic web environments in Chapter 4. If future spectroscopic surveys are to draw robust conclusions on the role different cosmic environments play on galaxy evolution, we must ask ourselves the question: can we reliably allocate individual galaxies to different cosmic web environments in the first place?

By taking into account the projection effects and fibre configuration technique that the WEAVE instrument will use, we assign individual galaxies to three different

types of environment: the cluster environment (C), filamentary environments (F) and the aptly named 'neither cluster nor filamentary environments' (NCF). After assigning galaxies to these three environments, we found that:

1. Filaments occupy only $\sim 6\%$ of the volume enclosed in a sphere with radius $5R_{200}$ around massive galaxy clusters, but contain $\sim 38\%$ of the galaxies with masses above $10^9 M_{\odot}$. This is calculated using filament thicknesses that decrease with the mass of the main halo (see Sec. 4.3.2). In comparison, galaxies that are neither in the cluster core nor in filaments (NCF) make up $\sim 52\%$ of the galaxy population, whilst occupying 93% of the volume. The cluster core itself (defined as the sphere with radius R_{200}) contains $\sim 1\%$ of the volume and $\sim 10\%$ 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.

2. 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% (68%), 63% (51%) and 62% (68%) respectively (see definitions in Equations 4.2 and 4.3). 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.

3. 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 one (Figure 4.4 and Table B.1). 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, and this also depends on the method used for

generating the random probabilities. 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).

4. 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 B.4, B.3, and B.2), 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.

In this Chapter 4 we conclude that in the infall regions of 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.

In addition, we have recently developed a 'proof of concept' framework which shows that we can use Machine Learning (ML) to improve upon our classifications published in Chapter 4. In Weaver et al. (2023), we train a Random Forest (RF; Ho 1995) classifier on information that is observational derivable. This study showed that, on average, we can improve the precision of filament classification by $\sim 11\%$. We discuss this work further in Section 6.2

6.1.3 Galaxy group detection in cluster outskirts

In Chapter 5, we investigate a new method to systematically detect galaxy groups in close proximity to galaxy clusters. Galaxy groups, (defined as objects with 1D velocity dispersion $\sigma_v > 300\ h^{-1}$ km/s, Kuchner et al. 2021), are abundant in the infall region of galaxy clusters and can influence galaxy evolution before galaxies are accreted onto the cluster (Bianconi et al., 2018). Therefore, they are

a region of significant astrophysical interest. As explained in Section 1.2.2, galaxy groups in the vicinity of massive clusters are difficult to detect. Currently, we are limited to laborious, non-systematic methods of group detection that are used on a cluster-by-cluster basis. Consequently, developing a tool to automatically detect these objects would be hugely beneficial.

In Chapter 3 & 4, we showed that we can accurately extract cosmic filament networks around galaxy clusters. In Chapter 5, to round off the motivation for the science carried out in this thesis, we test the use of the filament extractor `DisPerSE` in detecting galaxy groups. We have tested the matching of cosmic web nodes, to galaxy groups in and around massive galaxy clusters in 3D and projected 2D cluster simulations. We found that:

1. We return a purity of $\sim$ 56% and a completeness of $\sim$ 68% in 3D and a purity of $\sim$ 28% and a completeness of 26% in 2D in locating galaxy groups using cosmic web nodes.

2. Galaxy groups that closely match with cosmic web nodes tend to be the more massive ones.

3. In the 3D reference simulations, we found a slight improvement in the fractional number of nodes within galaxy groups as we move further away from the cluster core. This suggests that the cluster core's complexity hinders the accurate matching of density field maxima to galaxy groups. Conversely, in the cluster outskirts the success rate increases due to the dominant influence of the large scale cosmic web rather than the cluster core. Within the range of 3–5R_{200}, the number of nodes in groups reaches a maximum of approximately $\sim$ 75%, matching the results obtained by Cohn (2022) over much larger spatial scales. In contrast, in the 2D projections, we do not see any radial trend.

4. Limiting our analysis to only the most massive galaxy groups ($M > 10^{14} M_{\odot}$), we found that 100% of cosmic web nodes match to galaxy groups in the reference 3D simulations. We also found a stark improvement in the success rate for 2D projections, increasing from $\sim$ 28% to $\sim$ 63%.

5. We found a strong positive correlation between the mass of the groups and the `DisPerSE`-determined density of their matching nodes. This correlation

(equation 3) can be used to obtain a rough estimate of the group mass within a factor of ~ 3.

In summary, we have shown that the widely-used topological filament finder `DisPerSE` can be used as a useful tool for identifying galaxy groups around clusters. It can be further complimented with other group finding algorithms.

6.2 Future work

In the final section, we describe extensions to the work carried out in this thesis and their possible implications for future surveys.

6.2.1 Characterizing galaxy cluster mergers observationally

Galaxy clusters are not isolated islands, but assemble, replenish and grow via ongoing mergers with smaller clusters, groups and clumps of gas, as well as through a constant flow of gas and galaxies from filaments. As a result of the ongoing dynamical activity, the infall of matter is highly anisotropic meaning that clusters are far from being spherical (Carter & Metcalfe, 1980; Fabricant et al., 1984). Furthermore, there is evidence suggesting that galaxy cluster mergers can influence the morphologies and colours of galaxies residing in them (Contreras-Santos et al., 2022; Piraino-Cerda et al., 2023). Estimating a galaxy cluster's recent accretion history is therefore important for understanding the mass assembly of the clusters themselves and the evolution of cluster galaxies.

We are at quite an advanced stage in a recent project that aims to understand the success and limitations that upcoming wide-field spectroscopic surveys will have in mapping the recent dynamical history of a galaxy cluster. With the thousands of optical spectra these surveys will provide, we will gain extensive information about the dynamical properties of galaxies in galaxy clusters, (for example, the distributions of their positions, masses, velocities), as well as topological properties

of the local cosmic web (i.e. the distribution of cosmic web nodes and the paths the filaments trace).

Whilst it is fairly common practise to use the dynamics of galaxies in clusters to infer the dynamical state of a galaxy cluster, it is largely unknown whether the topology of the local cosmic web can be used to probe the cluster's mass accretion history. To motivate our use of the topological features of the cosmic web in our analysis, we refer to Chapter 5, where we cite the study of Cohn (2022). They found that clusters with a recent major merger are less likely to be identified as a node of the cosmic web. With this in mind, the basic idea of this project is to understand what combination of wide-field observable properties are the strongest indicators of the dynamical state of a cluster.

In this work, we use the outputs of the Multi-Dark Planck 2 (MDPL2) dark matter only N-body simulation (Klypin et al., 2016), which we describe in Section 2.1.1. We use the outputs of AHF ran on the 324 most massive haloes at $z = 0$. We also make use of the major merger history catalogues generated in Contreras-Santos et al. (2022) to divide the sample of 324 clusters into 52 merging clusters and 192 non-mergers at $z = 0$. The definition of merging and non-merging clusters we use in this work is based on the 'relaxation' ($\chi_{\rm DS}$) of a galaxy cluster, and is described in Haggar et al. (2020). It combines in quadrature the centre of mass offset (the offset between the centre of mass of the cluster and the density peak of the cluster halo, as a fraction of R_{200}), the subhalo mass fraction (the fraction of cluster mass contained in subhaloes) and the virial ratio (a measure of how well the cluster obeys the virial theorem). A merger finishes when the cluster has relaxed, i.e., when $\chi_{\rm DS} > 1$. In this work, we tag any cluster that is unrelaxed at $z = 0$ as a merger. We tag any cluster as a non-merger if it has not started a 1:2 merger in the last 5 Gyrs.

We have used a range of previously used and some novel parameters to understand the dynamical activity of a cluster. These include: the skewness (μ_3) and kurtosis (μ_4) of the radial velocity distribution, the triaxiality of the haloes in the cluster core (c/a and b/a, where a, b, c are the longest, second longest and third longest axis of the moment of inertia tensor), the distance between the main cluster halo and the second densest cosmic web node ($D_{\rm node}$) and finally, the ratio of density of the 'main' node to the second densest node $\frac{\delta_2}{\delta_1}$. We calculate the above parameters

in the 3D 'ground truth' galaxy cluster simulations and sufficiently adapt them where necessary in the mock 2D projections where we apply a cluster membership algorithm. We provide the details for all of this work in an upcoming paper, Cornwell et al. in prep., and discuss some of the main findings below.

Figure 6.1 illustrates the cumulative distributions of the third and fourth order moment of the radial velocity profiles. We perform a Kolmogorov-Smirnov test between the distribution of skewnesses for the merging and non-merging samples, (we also repeat this test separately using the kurtosis distributions). We test the null hypothesis that the skewness distribution of merging clusters could be drawn from that of the non-merging clusters and chose a significance threshold of 0.05. The top panel reveals that there is no statistical difference in the skewness between the merging and the non-merging sample. However, in the bottom panel, we show that the merging clusters have a more positive distribution of kurtosis on average, meaning that the radial velocity distributions are truncated and more peaked. We repeat the KS-test for the kurtosis of the velocity distributions. In 3D, we record a p-value of 0.003 and in the mock observations, a p-value of 0.024. This result is in line with Vijayaraghavan et al. (2015) who found that in a major cluster merger, during first pericentre passage, there is a spike in the kurtosis of the radial velocity distribution. This could be due to the more weakly bound haloes being stripped from the cluster they were initially bound to.

We repeat the same statistical test to probe the difference in the triaxiality of merging clusters from non-merging clusters. Crucially, when restricting haloes to R_{200}, the distribution of merging and non-merging clusters can be drawn from the same distribution, (p-value > 0.05). Therefore, we extend our test to include haloes within $2R_{200}$, where the dynamical timescale is approximately the same as our chosen merger/non-merger timescale cutoff of 5 Gyrs. We find a surprising result: merging clusters are more spherical and slightly more oblate than non-merging clusters. In the 2D projected clusters, a similar trend is seen, such that merging clusters are less elliptical at this radius. Both of these tests are significant given our chosen threshold, (p-value < 0.05). We hypothesize that the presence of a merger 'sweeps-up' the mass outside of the cluster core, resulting in a more uniform halo distribution outside of the cluster core. However, further simulation based work is needed to scrutinize the physics driving this result.

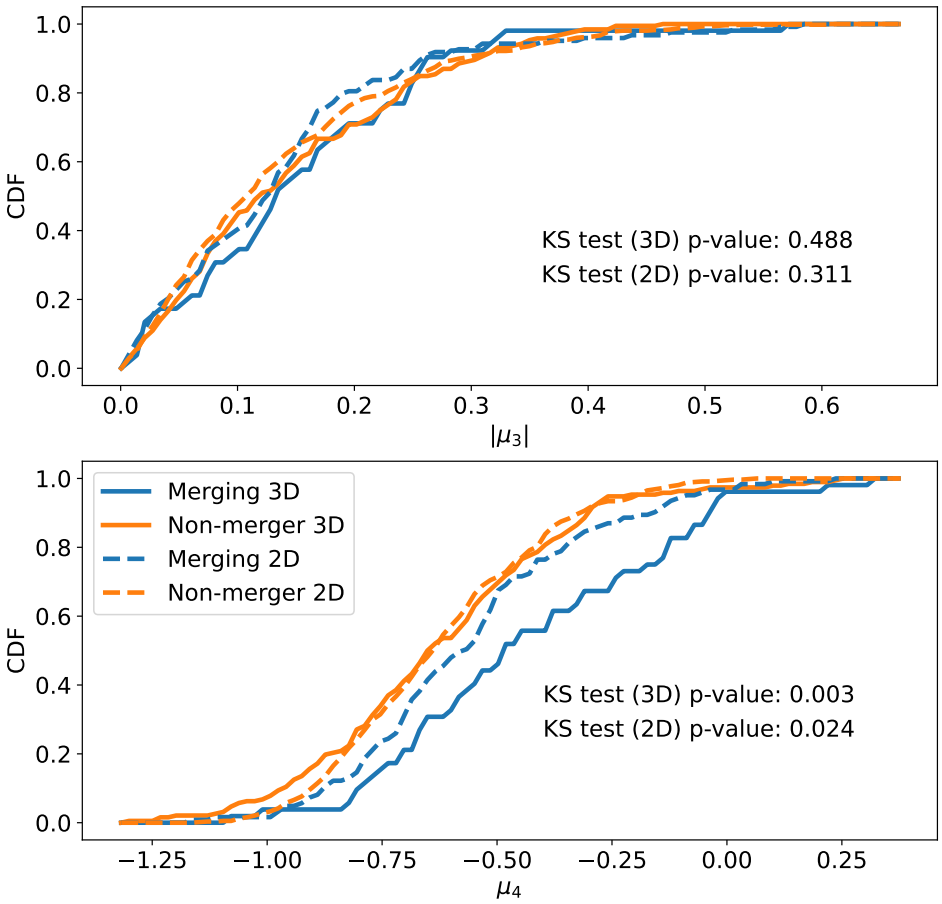


Figure 6.1: Third and fourth order moments (skewness and kurtosis) of the radial velocity distribution for merging and non-merging galaxy clusters. Top panel: A cumulative distribution function of the skewness μ_3 of the radial velocity distribution for merging and non merging clusters. In blue, we show the distributions of the merging clusters and in orange, we show the distributions of the non-merging clusters. We also show the distinction between the 3D and 2D clusters as full lines and dashed lines respectively. We display the p-value from the resulting KS test between the two distributions in the bottom of each plot for the 3D and 2D sample. Bottom panel: the same but for the kurtosis μ_4. Merging clusters have a larger kurtosis on average, indicating a compression of the line-of-sight radial distribution.

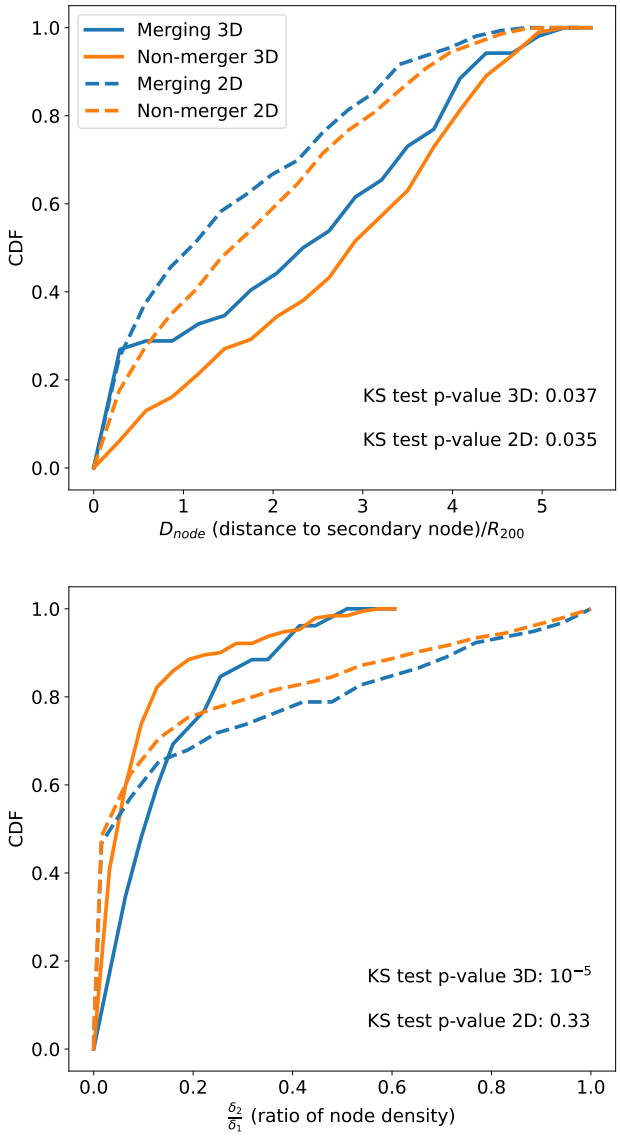


Figure 6.2: Topological metrics used to differentiate between merging and non-merging clusters. Top: CDF of the clustercentric distance to the second densest node $D_{\rm node}$, as computed by `DisPerSE`. In blue we show the merging clusters and in orange we show the non-merging clusters. Further to this, we illustrate the distributions as a full line in the 3D case and as a dashed line in the 2D case. Bottom: CDFs of the ratio of node densities $\frac{\delta_2}{\delta_1}$ for merging and non-merging clusters in 3D and 2D. We display the p-value from the resulting KS test between the two distributions in the bottom of each plot. Merging clusters have a shorter distance to the secondary node and a higher node density ratio, indicating a perturbation in the halo distribution in the cluster.

In Figure 6.2, we show the cumulative distribution functions (CDFs) of the two novel topological metrics we use to investigate the dynamical activity of a cluster. It is evident that in both 3D and 2D, $D_{\rm node}$ is smaller for merging clusters. In other words, merging clusters have a more prominent secondary node closer to the cluster core. This indicates that there is a disturbance to the underlying density field. In the bottom panel of Figure 6.2, we show the same test but using the node density ratio metric. The large offset between the merging and non-merging distributions in 3D indicate a more influential secondary node, agreeing with the result in the top panel. However, with the dimensional reduction implemented in 2D, we do not see the same signal. This implies that the ability of the Delaunay tessellation to reproduce the 'true' density field constructed in 3D is strongly impacted. Therefore, for the topological metrics we investigate, $D_{\rm node}$ appears to be a more robust indicator of dynamical state for future surveys.

To round off this project, we are now investigating what features most strongly indicate dynamical activity in clusters. To take this a step further, we are interested in understanding whether we can combine the different diagnostics we have tested to supply a pure sample of clusters that have recently undergone a major merger or are dynamically relaxed. This work could have both astrophysical and cosmological implications. In terms of galaxy evolution, with surveys like the WWFCS and CHANCES, it could allow investigations into the dependency of galaxy properties on a cluster's dynamical state. In a cosmology context, sets of dynamically relaxed clusters are required for testing models of dark energy (Mantz et al., 2010) and providing constraints on the gas mass fraction of clusters (Mantz et al., 2021). Along with other methods used to define the dynamical state of a cluster, it is possible that we will be able to derive a robust recent accretion history for these clusters, enabling the aforementioned studies.

6.2.2 Cosmic web environment classification with Machine Learning

In Chapter 4, we performed a statistical analysis that probed our ability to allocate mock galaxies to different cosmic web environments, namely the cluster core, filaments and galaxies that are in neither. One of the main conclusions from this

work was that assigning galaxies to filaments is incredibly difficult, due to the projection of the galaxies and the low contrast nature of filamentary environments. However, the salient point from this work was that for all types of galaxies, we can perform better than a random allocation, meaning that with large samples of galaxies, we will be able to beat the noise and draw robust conclusions on how filamentary environments impact galaxy evolution.

A recent work by Farid et al. (2023) showed that, using the MDPL2 simulations, it is possible to use Machine Learning (ML) to classify galaxies as orbiting, infalling and interlopers in the infall regions of massive galaxy clusters. Inspired by this work, we have recently embarked on a new project, where we are investigating whether the application of ML can improve upon the probabilities we calculated in Chapter 4, with the main aim of improving the filamentary environment classification.

In this project, we are initially using a *Random Forest* (RF; Ho 1995) classifier. Briefly, a RF classifier uses an ensemble of decision trees to group datasets into smaller subgroups that have similar features. It is a relatively simplistic ML algorithm that prevents over-fitting of data. For its input data, it uses a set of *elements*, each of which has a number of *features*, and is assigned to a *class*. For our work, an element is a galaxy around the simulated clusters, a feature is a property of these galaxies and the class is the environment classification it falls in to. The features we use are all observable parameters. A non-exhaustive feature list is: the mass and sky position of a galaxy, its radial velocity, the number of galaxies in a cluster and the distance to the nearest cosmic filament reconstructed in 2D. An extensive list of the features and choice of hyperparameters is provided in Weaver et al. (2023) and Weaver et al., in prep. A class is a binary classification, for example, a filament galaxy or a non-filament galaxy.

A RF works by firstly constructing a single decision tree that takes a random sub sample of elements and features[1]. The data is initially split into two subsets by slicing it as cleanly as possible using one of the features. Then, each subsequent set is split based on a different feature and the algorithm iterates through this process until a decision tree is constructed that separates the data into numerous subsets.

[1]the selection of a random sample of elements and features for each individual tree results in minimal overfitting.

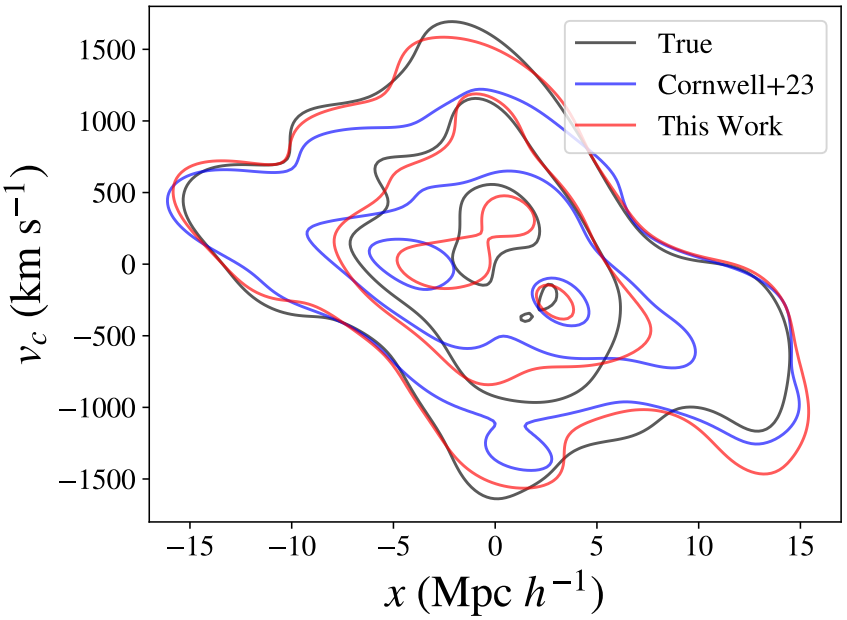


Figure 6.3: Distribution of $x - v_c$ projected phase space of haloes classified as filamentary for a single cluster. The reconstruction of the parameters plotted in this phase space diagram better match the ground truth. This illustrates that RF classification is more successful in classifying filament galaxies than the work we carried out in Chapter 4. Figure adapted from Weaver et al. (2023).

In our case, the tree ceases to bifurcate when an enforced maximum tree depth is reached and, consequently, each subset mostly contains elements of a single class. Secondly, the RF classifier repeats this whole process using a subset of elements, reconstructing a large number of decision trees that samples the entire training data and features. Finally, these trees are combined to act as an ensemble, and are averaged over to enable predictions of which class an element will fall into. In turn, the hyper-parameters of the RF algorithm can be fine tuned, allowing an optimization of the RF network.

We have recently published a 'proof of concept' research note where we present our preliminary results (Weaver et al., 2023). Using a RF classifier, we can improve the precision of filament galaxy allocation by over 11% compared to the results presented in Chapter 4. Figure 6.3 shows the x - v_c projected phase space (PPS) for a single cluster, (where x and v_c are the x position and the line of sight velocity of a galaxy that is classified as a filament galaxy). In black, it shows the galaxies that are 'truly' filamentary, in blue it shows the galaxies assigned as filamentary

using the methodology we explained in Chapter 4 and in red, the results from the RF classification used in Weaver et al. (2023). For this example cluster, it is clear that we can more accurately reproduce the 'true' PPS using a RF classifier compared to the work presented in Chapter 4. These preliminary results show that ML can be used as a powerful tool in identifying filamentary galaxies.

The next steps in this project are to firstly understand what combination of features returns the classification with the highest precision and accuracy. Secondly, we will fine-tune the hyperparameters with the aim of increasing the number of successful classifications. Lastly, as in Farid et al. (2023), we are also investigating whether other ML classifiers can help us improve upon these results. With surveys like the WWFCS and CHANCES on the horizon, these findings are essential. By driving an increase in the accuracies and precisions in galaxy environment classification, we will be able to draw more robust conclusions on how galaxy evolution is shaped by low contrast environments – a key motivation of these surveys.

6.3 Upcoming observations

The next several years will bring about a variety of rich, detailed observational surveys of galaxy clusters. These observations include a range of high quality optical spectroscopic surveys and deep X-ray observations, covering the cluster core and associated substructure, such as cosmic filaments and infalling galaxy groups. Below, we name a few examples of exciting next generation surveys that may aim to utilize and build on the work carried out in this thesis.

Whilst we have talked extensively about the WEAVE Wide Field Cluster Survey, an upcoming survey with similar scientific aims is the 4MOST CHileAN Cluster galaxy Evolution Survey (CHANCES; Haines et al., 2023). This survey will utilize a multi-object spectrograph fit to the VLT and target 150 galaxy clusters, out to $z = 0.45$, enabling spectroscopy of thousands of galaxies per cluster out to $5R_{200}$. Similarly to the WWFCS, CHANCES will probe down the galaxy stellar mass function to $\approx 10^9 M_{\odot}$, aiming to quantify the impact of environment on galaxy evolution across three orders of magnitude in stellar mass. Given the larger redshift range than the WWFCS, CHANCES will also aim to probe the evolution of cluster

galaxies over the last 4 Gyrs. This survey will be able to infer where and when pre-processing occurs as the cluster itself is evolving. Finally, with complimentary X-ray observations of the ICM with the CHEX-MATE survey (Arnaud et al., 2021), CHANCES will robustly identify how the dynamical state of a galaxy cluster may influence the galaxies residing in it.

The work carried out in this thesis will compliment the WWFCS/CHANCES perfectly. The framework we have described in Chapter 3, 4 & 6.2.2 demonstrated the relative success and limitations in cosmic web environment classification. Taking into account the proper statistical treatment that we developed will enable the aforementioned surveys to draw robust and well quantified conclusions on the role that the cosmic web plays in the mass assembly and evolution of individual galaxies. Furthermore, given CHANCES complimentary X-ray observations, the merger history diagnostics we are developing in Section 6.2.1 can be validated against the more commonly used X-ray measurements with CHEX-MATE. They could also be potentially combined to provide a more robust description of the accretion history over a range of timescales.

In addition to these ground based spectroscopic surveys that will sample a relatively small number of clusters, *Euclid* (Sartoris et al., 2016) will provide imaging for $> 10^5$ clusters. Whilst the target densities are expected to be lower than the WWFCS/CHANCES, such a large sample of clusters will provide excellent cluster statistics. Also, it will produce rigorous constraints on cosmological parameters and the dynamical evolution of dark energy. Finally, in the next decade, the MOSAIC spectroscopic instrument will be fit to the ELT and commissioned (Jagourel et al., 2018). This survey will be able to probe how galaxies are shaped by their cosmic environment back to the epoch of the peak of cosmic star formation at $z \approx 2$ (Madau & Dickinson, 2014).

The near future of cosmic web science, along with the astrophysics of galaxy clusters and galaxy evolution, is incredibly exiting. By merging insights from upcoming observations with cutting-edge simulations, we will be able to draw robust conclusions on how galaxies are influenced by their specific cosmic web environments. This knowledge will significantly enhance our understanding of the complex journeys galaxies undergo during their formation and evolution. The research presented in this thesis contributes to these advancements by developing frame-

works that enable the observational characterization of galaxies in low-density environments.

Bibliography

Abell, G. O. 1957, The Astronomical Journal, 62, 2, doi: 10.1086/107564

Alpaslan, M., Grootes, M., Marcum, P. M., et al. 2016, Monthly Notices of the Royal Astronomical Society, 457, 2287–2300, doi: 10.1093/mnras/stw134

Angulo, R. E., & Hahn, O. 2022, Living Reviews in Computational Astrophysics, 8, doi: 10.1007/s41115-021-00013-z

Aragón-Calvo, M. A., van de Weygaert, R., & Jones, B. J. T. 2010, Monthly Notices of the Royal Astronomical Society, 408, 2163, doi: 10.1111/j.1365-2966.2010.17263.x

Arnaud, M., Ettori, S., Pratt, G. W., et al. 2021, A&A, 650, A104, doi: 10.1051/0004-6361/202039632

Arnold, V. I., Shandarin, S. F., & Zeldovich, Y. B. 1982, Geophysical & Astrophysical Fluid Dynamics, 20, 111, doi: 10.1080/03091928208209001

Arthur, J., Pearce, F. R., Gray, M. E., et al. 2019, Monthly Notices of the Royal Astronomical Society, 484, 3968, doi: 10.1093/mnras/stz212

Bahé, Y. M., & McCarthy, I. G. 2014, Monthly Notices of the Royal Astronomical Society, 447, 969–992, doi: 10.1093/mnras/stu2293

Balcells, M., Benn, C. R., Carter, D., et al. 2010, in Ground-based and Airborne Instrumentation for Astronomy III, Vol. 7735, International Society for Optics and Photonics, 2659–2673

Balogh, M. L., Gilbank, D. G., Muzzin, A., et al. 2017, Monthly Notices of the Royal Astronomical Society, 470, 4168, doi: 10.1093/mnras/stx1370

Beck, A. M., Murante, G., Arth, A., et al. 2015, Monthly Notices of the Royal Astronomical Society, 455, 2110, doi: 10.1093/mnras/stv2443

Beck, R., Szapudi, I., Flewelling, H., et al. 2020, Monthly Notices of the Royal Astronomical Society, 500, 1633, doi: 10.1093/mnras/staa2587

Benavides, J. A., Sales, L. V., & Abadi, M. G. 2020, Monthly Notices of the Royal Astronomical Society, 498, 3852, doi: 10.1093/mnras/staa2636

Berlind, A. A., Frieman, J., Weinberg, D. H., et al. 2006, The Astrophysical Journal Supplement Series, 167, 1, doi: 10.1086/508170

Bianconi, M., Smith, G. P., Haines, C. P., et al. 2017, Monthly Notices of the Royal Astronomical Society: Letters, 473, L79, doi: 10.1093/mnrasl/slx167

Bianconi, M., Smith, G. P., Haines, C. P., et al. 2018, Monthly Notices of the Royal Astronomical Society, 473, L79, doi: 10.1093/mnrasl/slx167

Bond, J, R., Kofman, L., & Pogosyan, D. 1996, Nature, 380, 603

Broadhurst, T. J., Taylor, A. N., & Peacock, J. A. 1995, The Astrophysical Journal, 438, 49, doi: 10.1086/175053

Burchett, J. N., Elek, O., Tejos, N., et al. 2020, The Astrophysical Journal Letters, 891, L35, doi: 10.3847/2041-8213/ab700c

Carter, D., & Metcalfe, N. 1980, Monthly Notices of the Royal Astronomical Society, 191, 325, doi: 10.1093/mnras/191.2.325

Castignani, G., Combes, F., Jablonka, P., et al. 2021, Astronomy & Astrophysics, 657, A9, doi: 10.1051/0004-6361/202040141

Castignani, G., Vulcani, B., Finn, R. A., et al. 2022, The Astrophysical Journal Supplement Series, 259, 43, doi: 10.3847/1538-4365/ac45f7

Cautun, M., van de Weygaert, R., Jones, B. J. T., & Frenk, C. S. 2014, Monthly Notices of the Royal Astronomical Society, 441, 2923–2973, doi: 10.1093/mnras/stu768

Cautun, M. C., & van de Weygaert, R. 2019, The DTFE public software: The Delaunay Tessellation Field Estimator code. https://arxiv.org/abs/1105.0370

Cenarro, A. J., Moles, M., Cristóbal-Hornillos, D., et al. 2019, Astronomy & Astrophysics, 622, A176, doi: 10.1051/0004-6361/201833036

Chen, Y.-C., et al. 2017, MNRAS, 466, 1880–1893

Cohn, J. D. 2022, Monthly Notices of the Royal Astronomical Society, 513, 624, doi: 10.1093/mnras/stac894

Colless, M., Dalton, G., Maddox, S., et al. 2001, Monthly Notices of the Royal Astronomical Society, 328, 1039, doi: 10.1046/j.1365-8711.2001.04902.x

Contreras-Santos, A., Knebe, A., Pearce, F., et al. 2022, Monthly Notices of the Royal Astronomical Society, 511, 2897, doi: 10.1093/mnras/stac275

Cornwell, D. J., Aragón-Salamanca, A., Kuchner, U., et al. 2023a, Monthly Notices of the Royal Astronomical Society, stad1949, doi: 10.1093/mnras/stad1949

Cornwell, D. J., Kuchner, U., Gray, M. E., et al. 2023b, Monthly Notices of the Royal Astronomical Society, stad3205, doi: 10.1093/mnras/stad3205

Cornwell, D. J., Kuchner, U., Aragón-Salamanca, A., et al. 2022, Monthly Notices of the Royal Astronomical Society, 517, 1678, doi: 10.1093/mnras/stac2777

Cortese, L., Gavazzi, G., Boselli, A., et al. 2006, A&A, 453, 847–861, doi: 10.1051/0004-6361:20064873

Cui, W., Knebe, A., Yepes, G., et al. 2018, Monthly Notices of the Royal Astronomical Society, 480, 2898, doi: 10.1093/mnras/sty2111

Cui, W., Dave, R., Knebe, A., et al. 2022, Monthly Notices of the Royal Astronomical Society, doi: 10.1093/mnras/stac1402

Dalton, G. 2016, in Astronomical Society of the Pacific Conference Series, Vol. 507, Multi-Object Spectroscopy in the Next Decade: Big Questions, Large Surveys, and Wide Fields, ed. I. Skillen, M. Balcells, & S. Trager, 97

Dalton, G., Trager, S., Abrams, D. C., et al. 2014, in Ground-based and airborne instrumentation for astronomy V, Vol. 9147, SPIE, 177–187

Darragh Ford, E., Laigle, C., Gozaliasl, G., et al. 2019, Monthly Notices of the Royal Astronomical Society, 489, 5695, doi: 10.1093/mnras/stz2490

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, The Astrophysical Journal, 292, 371, doi: 10.1086/163168

Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, Monthly Notices of the Royal Astronomical Society, 486, 2827, doi: 10.1093/mnras/stz937

de Lapparent, V., Geller, M. J., & Huchra, J. P. 1989, The Astrophysical Journal, 343, 1, doi: 10.1086/167679

De Luca, F., De Petris, M., Yepes, G., et al. 2021, Monthly Notices of the Royal Astronomical Society, 504, 5383, doi: 10.1093/mnras/stab1073

Di Mascolo, L., Saro, A., Mroczkowski, T., et al. 2023, Nature, 615, 809–812, doi: 10.1038/s41586-023-05761-x

Diaferio, A., & Geller, M. J. 1997, The Astrophysical Journal, 481, 633–643, doi: 10.1086/304075

Dolag, K., Meneghetti, M., Moscardini, L., Rasia, E., & Bonaldi, A. 2006, Monthly Notices of the Royal Astronomical Society, 370, 656, doi: 10.1111/j.1365-2966.2006.10511.x

Dressler, A. 1980, The Astrophysical Journal, 236, 351, doi: 10.1086/157753

Dressler, A., & Shectman, S. A. 1988, The Astrophysical Journal, 95, 985, doi: 10.1086/114694

Dressler, A., Oemler, Augustus, J., Couch, W. J., et al. 1997, The Astrophysical Journal, 490, 577, doi: 10.1086/304890

Einasto, M., Deshev, Boris, Tenjes, Peeter, et al. 2020, A&A, 641, A172, doi: 10.1051/0004-6361/202037982

Eke, V. R., Frenk, C. S., Baugh, C. M., et al. 2004, Monthly Notices of the Royal Astronomical Society, 355, 769, doi: 10.1111/j.1365-2966.2004.08354.x

Evrard, A. E. 1990, The Astrophysical Journal, 363, 349, doi: 10.1086/169350

Fabricant, D., Rybicki, G., & Gorenstein, P. 1984, The Astrophysical Journal, 286, 186, doi: 10.1086/162586

Farid, D., Aung, H., Nagai, D., Farahi, A., & Rozo, E. 2023, Astronomy and Computing, 45, 100743, doi: 10.1016/j.ascom.2023.100743

Fasano, G., Marmo, C., Varela, J., et al. 2006, A&A, 445, 805, doi: 10.1051/0004-6361:20053816

Finn, R. A., Zaritsky, D., Donald W. McCarthy, J., et al. 2005, The Astrophysical Journal, 630, 206, doi: 10.1086/431642

Finoguenov, A., Briel, U. G., & Henry, J. P. 2003, A&A, 410, 777–784, doi: 10.1051/0004-6361:20031319

Galárraga-Espinosa, Langer, Mathieu, & Aghanim, Nabila. 2022, A&A, 661, A115, doi: 10.1051/0004-6361/202141974

Galárraga-Espinosa, D., Cadiou, C., Gouin, C., et al. 2024, A&A, 684, A63, doi: 10.1051/0004-6361/202347982

Ganeshaiah Veena, P., Cautun, M., van de Weygaert, R., et al. 2018, Monthly Notices of the Royal Astronomical Society, 481, 414, doi: 10.1093/mnras/sty2270

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

Gladders, M. D., & Yee, H. K. C. 2000, The Astronomical Journal, 120, 2148–2162, doi: 10.1086/301557

Gonzalez, A. H., Sivanandam, S., Zabludoff, A. I., & Zaritsky, D. 2013, The Astrophysical Journal, 778, 14, doi: 10.1088/0004-637x/778/1/14

Gouin, C., Aghanim, N., Bonjean, V., & Douspis, M. 2020, Astronomy & Astrophysics, 635, A195, doi: 10.1051/0004-6361/201937218

Gouin, C., Bonnaire, T., & Aghanim, N. 2021, A&A, 651, A56, doi: 10.1051/0004-6361/202140327

Gouin, C., Gallo, S., & Aghanim, N. 2022, A&A. http://arxiv.org/abs/2201.00593

Gray, M. E., Wolf, C., Barden, M., et al. 2009, Monthly Notices of the Royal Astronomical Society, 393, 1275, doi: 10.1111/j.1365-2966.2008.14259.x

Gullieuszik, M., Poggianti, B., Fasano, G., et al. 2015, Astronomy & Astrophysics, 581, A41, doi: 10.1051/0004-6361/201526061

Gunn, J. E., & Gott, J. Richard, I. 1972, The Astrophysical Journal, 176, 1, doi: 10.1086/151605

Haggar, R., Gray, M. E., Pearce, F. R., et al. 2020, Monthly Notices of the Royal Astronomical Society, 492, 6074, doi: 10.1093/mnras/staa273

Haines, C., Jaffé, Y., Tejos, N., et al. 2023, The Messenger, 190, 31

Haines, C. P., et al. 2015, ApJ, 806, 101,123

Haines, C. P., Finoguenov, A., Smith, G. P., et al. 2018, Monthly Notices of the Royal Astronomical Society, 477, 4931–4950, doi: 10.1093/mnras/sty651

Hickson, P. 1982, The Astrophysical Journal, 255, 382, doi: 10.1086/159838

Hickson, P., Mendes de Oliveira, C., Huchra, J. P., & Palumbo, G. G. 1992, The Astrophysical Journal, 399, 353, doi: 10.1086/171932

Ho, T. K. 1995, in Proceedings of 3rd International Conference on Document Analysis and Recognition, Vol. 1, 278–282 vol.1, doi: 10.1109/ICDAR.1995.598994

Hoekstra, H., Bartelmann, M., Dahle, H., et al. 2013, Space Science Reviews, 177, 75–118, doi: 10.1007/s11214-013-9978-5

Hoosain, M., Blyth, S.-L., Skelton, R. E., et al. 2024, The effect of cosmic web filaments on galaxy properties in the RESOLVE and ECO surveys. https://arxiv.org/abs/2401.09114

Hopkins, P. F. 2015, Monthly Notices of the Royal Astronomical Society, 450, 53–110, doi: 10.1093/mnras/stv195

Huchra, J., Davis, M., Latham, D., & Tonry, J. 1983, The Astrophysical Journals, 52, 89, doi: 10.1086/190860

Huchra, J. P., & Geller, M. J. 1982, The Astrophysical Journal, 257, 423, doi: 10.1086/160000

Hughes, S., Dalton, G., Dee, K., et al. 2022, in Ground-based and Airborne Instrumentation for Astronomy IX, ed. C. J. Evans, J. J. Bryant, & K. Motohara (SPIE), doi: 10.1117/12.2627132

Jackson, J. M., Finn, S. C., Chambers, E. T., Rathborne, J. M., & Simon, R. 2010, The Astrophysical Journal Letters, 719, L185, doi: 10.1088/2041-8205/719/2/L185

Jaffé, Y. L., Poggianti, B. M., Verheijen, M. A. W., Deshev, B. Z., & van Gorkom, J. H. 2013, Monthly Notices of the Royal Astronomical Society, 431, 2111, doi: 10.1093/mnras/stt250

Jaffé, Y. L., Verheijen, M. A. W., Haines, C. P., et al. 2016, Monthly Notices of the Royal Astronomical Society, 461, 1202–1221, doi: 10.1093/mnras/stw984

Jagourel, P., Fitzsimons, E., Hammer, F., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10702, Ground-based and Airborne Instrumentation for Astronomy VII, ed. C. J. Evans, L. Simard, & H. Takami, 10702A4, doi: 10.1117/12.2314135

Jin, S., Trager, S. C., Dalton, G. B., et al. 2023, Monthly Notices of the Royal Astronomical Society, 530, 2688–2730, doi: 10.1093/mnras/stad557

Jung, S. L., Choi, H., Wong, O. I., et al. 2018, The Astrophysical Journal, 865, 156, doi: 10.3847/1538-4357/aadda2

Kaiser, N. 1986, Monthly Notices of the Royal Astronomical Society, 222, 323, doi: 10.1093/mnras/222.2.323

Kirkpatrick S., Gelatt C. D., V. M. P. 1983, Sci, 220, 671

Kleiner, D., Pimbblet, K. A., Heath Jones, D., Koribalski, B. S., & Serra, P. 2016, Monthly Notices of the Royal Astronomical Society, stw3328, doi: 10.1093/mnras/stw3328

Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, Monthly Notices of the Royal Astronomical Society, 457, 4340, doi: 10.1093/mnras/stw248

Knebe, A., Knollmann, S. R., Muldrew, S. I., et al. 2011, Monthly Notices of the Royal Astronomical Society, 415, 2293, doi: 10.1111/j.1365-2966.2011.18858.x

Kraljic, K., Arnouts, S., Pichon, C., et al. 2017, Monthly Notices of the Royal Astronomical Society, 474, 547, doi: 10.1093/mnras/stx2638

Kraljic, K., Laigle, C., Pichon, C., et al. 2022, Monthly Notices of the Royal Astronomical Society, 514, 1359, doi: 10.1093/mnras/stac1409

Kuchner, U., Ziegler, B., Verdugo, M., Bamford, S., & Häußler, B. 2017, A$A, 604, A54, doi: 10.1051/0004-6361/201630252

Kuchner, U., Aragón-Salamanca, A., Pearce, F. R., et al. 2020, Monthly Notices of the Royal Astronomical Society, 494, 5473, doi: 10.1093/mnras/staa1083

Kuchner, U., Aragón-Salamanca, A., Rost, A., et al. 2021, Monthly Notices of the Royal Astronomical Society, 503, 2065, doi: 10.1093/mnras/stab567

Kuchner, U., Haggar, R., Aragón-Salamanca, A., et al. 2022, Monthly Notices of the Royal Astronomical Society, 510, 581, doi: 10.1093/mnras/stab3419

Kuzio de Naray, R., McGaugh, S. S., & Mihos, J. C. 2009, The Astrophysical Journal, 692, 1321–1332, doi: 10.1088/0004-637x/692/2/1321

Laigle, C., Pichon, C., Arnouts, S., et al. 2017, Monthly Notices of the Royal Astronomical Society, 474, 5437, doi: 10.1093/mnras/stx3055

Larson, R. B., Tinsley, B. M., & Caldwell, C. N. 1980, ApJ, 237, 692

Lau, E. T., Kravtsov, A. V., & Nagai, D. 2009, The Astrophysical Journal, 705, 1129–1138, doi: 10.1088/0004-637x/705/2/1129

Le Brun, A. M. C., McCarthy, I. G., Schaye, J., & Ponman, T. J. 2014, Monthly Notices of the Royal Astronomical Society, 441, 1270–1290, doi: 10.1093/mnras/stu608

Libeskind, N. I., van de Weygaert, R., Cautun, M., et al. 2017, Monthly Notices of the Royal Astronomical Society, 473, 1195, doi: 10.1093/mnras/stx1976

Lin, Y.-T., & Mohr, J. J. 2004, The Astrophysical Journal, 617, 879, doi: 10.1086/425412

Liske, J., Baldry, I. K., Driver, S. P., et al. 2015, Monthly Notices of the Royal Astronomical Society, 452, 2087, doi: 10.1093/mnras/stv1436

Lopes, P. A. A., Ribeiro, A. L. B., & Brambila, D. 2023, Monthly Notices of the Royal Astronomical Society: Letters, 527, L19–L25, doi: 10.1093/mnrasl/slad134

Lovisari, L., Ettori, S., Gaspari, M., & Giles, P. A. 2021, Universe, 7, 139, doi: 10.3390/universe7050139

Madau, P., & Dickinson, M. 2014, Annual Review of Astronomy and Astrophysics, 52, 415–486, doi: 10.1146/annurev-astro-081811-125615

Malavasi, N., Aghanim, Nabila, Tanimura, Hideki, Bonjean, Victor, & Douspis, Marian. 2020, A&A, 634, A30, doi: 10.1051/0004-6361/201936629

Malavasi, N., Langer, M., Aghanim, N., Galárraga-Espinosa, D., & Gouin, C. 2022, A&A, 658, A113, doi: 10.1051/0004-6361/202141723

Malavasi, N., Arnouts, S., Vibert, D., et al. 2016, Monthly Notices of the Royal Astronomical Society, 465, 3817, doi: 10.1093/mnras/stw2864

Malavasi, Nicola, Aghanim, Nabila, Tanimura, Hideki, Bonjean, Victor, & Douspis, Marian. 2020, A&A, 634, A30, doi: 10.1051/0004-6361/201936629

Mantz, A., Allen, S. W., Rapetti, D., & Ebeling, H. 2010, Monthly Notices of the Royal Astronomical Society, no, doi: 10.1111/j.1365-2966.2010.16992.x

Mantz, A. B., Morris, R. G., Allen, S. W., et al. 2021, Monthly Notices of the Royal Astronomical Society, 510, 131–145, doi: 10.1093/mnras/stab3390

Martel, H., Robichaud, F., & Barai, P. 2014, The Astrophysical Journal, 786, 79, doi: 10.1088/0004-637x/786/2/79

Martínez, H. J., Muriel, H., & Coenda, V. 2015, Monthly Notices of the Royal Astronomical Society, 455, 127, doi: 10.1093/mnras/stv2295

Martizzi, D., Vogelsberger, M., Artale, M. C., et al. 2019, Monthly Notices of the Royal Astronomical Society, 486, 3766–3787, doi: 10.1093/mnras/stz1106

McGee, S. L., Balogh, M. L., Bower, R. G., Font, A. S., & McCarthy, I. G. 2009, Monthly Notices of the Royal Astronomical Society, 400, 937–950, doi: 10.1111/j.1365-2966.2009.15507.x

Merritt, D. 1984, The Astrophysical Journal, 276, 26, doi: 10.1086/161590

Mohammad, F. G., Granett, B. R., Guzzo, L., et al. 2018, A&A, 610, A59, doi: 10.1051/0004-6361/201731685

Mohr, J. J., Mathiesen, B., & Evrard, A. E. 1999, The Astrophysical Journal, 517, 627–649, doi: 10.1086/307227

Moore, B., Lake, G., & Katz, N. 1998, The Astrophysical Journal, 495, 139, doi: 10.1086/305264

Morell, D. F., Ribeiro, A. L. B., de Carvalho, R. R., et al. 2020, Monthly Notices of the Royal Astronomical Society, 494, 3317, doi: 10.1093/mnras/staa881

Moretti, A., Gullieuszik, M., Poggianti, B., et al. 2017, A&A, 599, A81, doi: 10.1051/0004-6361/201630030

Mostoghiu, R., Knebe, A., Cui, W., et al. 2018, Monthly Notices of the Royal Astronomical Society, 483, 3390, doi: 10.1093/mnras/sty3306

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, The Astrophysical Journal, 490, 493, doi: 10.1086/304888

Okabe, N., Smith, G. P., Umetsu, K., Takada, M., & Futamase, T. 2013, The Astrophysical Journal, 769, L35, doi: 10.1088/2041-8205/769/2/l35

Ostriker, J. P., & Hausman, M. A. 1977, The Astrophysical Journal Letters, 217, L125, doi: 10.1086/182554

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, The Astrophysical Journal, 517, 565, doi: 10.1086/307221

Peterson, J., & Fabian, A. 2006, Physics Reports, 427, 1–39, doi: 10.1016/j.physrep.2005.12.007

Piraino-Cerda, F., Jaffé, Y. L., Lourenço, A. C., et al. 2023, Monthly Notices of the Royal Astronomical Society, 528, 919–936, doi: 10.1093/mnras/stad3957

Poggianti, B. M., De Lucia, G., Varela, J., et al. 2010, Monthly Notices of the Royal Astronomical Society, 405, 995, doi: 10.1111/j.1365-2966.2010.16546.x

Ponman, T. J., & Bertram, D. 1993, Nature, 363, 51, doi: 10.1038/363051a0

Poudel, A., Heinämäki, P., Tempel, E., et al. 2017, A&A, 597, A86, doi: 10.1051/0004-6361/201629639

Rasia, E., Borgani, S., Murante, G., et al. 2015, The Astrophysical Journal, 813, L17, doi: 10.1088/2041-8205/813/1/l17

Ribeiro, A. L. B., Lopes, P. A. A., & Rembold, S. B. 2013, A&A, 556, A74, doi: 10.1051/0004-6361/201220801

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, The Astrophysical Journal, 116, 1009, doi: 10.1086/300499

Robotham, A. S. G., Norberg, P., Driver, S. P., et al. 2011, Monthly Notices of the Royal Astronomical Society, 416, 2640, doi: 10.1111/j.1365-2966.2011.19217.x

Rost, A., Kuchner, U., Welker, C., et al. 2021, Monthly Notices of the Royal Astronomical Society, 502, 714, doi: 10.1093/mnras/staa3792

Salerno, J. M., Martínez, H. J., Muriel, H., et al. 2020, Monthly Notices of the Royal Astronomical Society, 493, 4950, doi: 10.1093/mnras/staa545

Sarazin, C. L. 1986, Reviews of Modern Physics, 58, 1, doi: 10.1103/RevModPhys.58.1

Sarron, F., Adami, C., Durret, F., & Laigle, C. 2019, Astronomy and Astrophysics, 632, doi: 10.1051/0004-6361/201935394

Sartoris, B., Biviano, A., Fedeli, C., et al. 2016, Monthly Notices of the Royal Astronomical Society, 459, 1764–1780, doi: 10.1093/mnras/stw630

Scarlata, C., Siana, B. D., Richard, J., et al. 2013, in American Astronomical Society Meeting Abstracts, Vol. 221, American Astronomical Society Meeting Abstracts #221, 210.03

Schaap, W. E., & van de Weygaert, R. 2000, Astronomy & Astrophysics, 363, L29, doi: 10.48550/arXiv.astro-ph/0011007

Secco, L. F., Samuroff, S., Krause, E., et al. 2022, Phys. Rev. D, 105, 023515, doi: 10.1103/PhysRevD.105.023515

Sembolini, F., Yepes, G., De Petris, M., et al. 2012, Monthly Notices of the Royal Astronomical Society, 429, 323, doi: 10.1093/mnras/sts339

Shandarin, S. F., & Zeldovich, Y. B. 1989, Rev. Mod. Phys., 61, 185, doi: 10.1103/RevModPhys.61.185

Song, H., Laigle, C., Hwang, H. S., et al. 2020, Monthly Notices of the Royal Astronomical Society, 501, 4635, doi: 10.1093/mnras/staa3981

Soucail, G., Mellier, Y., Fort, B., Mathez, G., & Cailloux, M. 1988, Astronomy & Astrophysics, 191, L19

Sousbie, T. 2011, Monthly Notices of the Royal Astronomical Society, 414, 350, doi: 10.1111/j.1365-2966.2011.18394.x

Sousbie, T., Pichon, C., & Kawahara, H. 2011, Monthly Notices of the Royal Astronomical Society, 414, 384, doi: 10.1111/j.1365-2966.2011.18395.x

Springel, V. 2005, Monthly Notices of the Royal Astronomical Society, 364, 1105, doi: 10.1111/j.1365-2966.2005.09655.x

Springel, V., Frenk, C. S., & White, S. D. M. 2006, Nature, 440, 1137, doi: 10.1038/nature04805

Springel, V., Yoshida, N., & White, S. D. 2001, New Astronomy, 6, 79–117, doi: 10.1016/s1384-1076(01)00042-2

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629, doi: 10.1038/nature03597

Stein, E., Milnor, J. W., Spivak, M., et al. 1963, Morse Theory (Princeton University Press)

Stephan, E. 1881, Astronomische Nachrichten, 100, 209, doi: 10.1002/asna.18811001402

Sunyaev, R. A., & Zeldovich, Y. B. 1972, Comments on Astrophysics and Space Physics, 4, 173

Tanimura, H., Aghanim, N., Kolodzig, A., Douspis, M., & Malavasi, N. 2020, A&A, 643, L2, doi: 10.1051/0004-6361/202038521

Taverna, A., Salerno, J. M., Daza-Perilla, I. V., et al. 2023, Monthly Notices of the Royal Astronomical Society, 520, 6367–6381, doi: 10.1093/mnras/stad416

Tawfeek, A. A., Sodi, B. C., Fritz, J., et al. 2022, The Astrophysical Journal, 940, 1, doi: 10.3847/1538-4357/ac9976

Tempel, E., Guo, Q., Kipper, R., & Libeskind, N. I. 2015, Monthly Notices of the Royal Astronomical Society, 450, 2727–2738, doi: 10.1093/mnras/stv919

Tempel, E., Kipper, R., Tamm, A., et al. 2016, A&A, 588, A14, doi: 10.1051/0004-6361/201527755

Tero, A., Takagi, S., Saigusa, T., et al. 2010, Science, 327, 439, doi: 10.1126/science.1177894

Terrett, D. L., Lewis, I. J., Dalton, G., et al. 2014, SPIE, 9152, 216, doi: 10.1117/12.2055844

Tully, R. B., & Fisher, J. R. 1978, The Large Scale Structure of the Universe; Proceedings of the Symposium, 31

Vallés-Pérez, D., Planelles, S., & Quilis, V. 2020, Monthly Notices of the Royal Astronomical Society, 499, 2303, doi: 10.1093/mnras/staa3035

Vazza, F., & Feletti, A. 2020, Frontiers in Physics, 8, 491, doi: 10.3389/fphy.2020.525731

Vernstrom, T., Heald, G., Vazza, F., et al. 2021, Monthly Notices of the Royal Astronomical Society, 505, 4178, doi: 10.1093/mnras/stab1301

Vijayaraghavan, R., Gallagher, J. S., & Ricker, P. M. 2015, Monthly Notices of the Royal Astronomical Society, 447, 3623, doi: 10.1093/mnras/stu2761

Vijayaraghavan, R., & Ricker, P. M. 2013, Monthly Notices of the Royal Astronomical Society, 435, 2713–2735, doi: 10.1093/mnras/stt1485

Vulcani, B., Poggianti, B. M., Moretti, A., et al. 2019, Monthly Notices of the Royal Astronomical Society, 487, 2278, doi: 10.1093/mnras/stz1399

Walker, S., & Lau, E. 2022, Cluster outskirts and their connection to the cosmic web, arXiv, doi: 10.48550/ARXIV.2202.07056

Weaver, T. S., Aung, H., Cornwell, D. J., Nagai, D., & Aragón-Salamanca, A. 2023, Research Notes of the American Astronomical Society, 7, 268, doi: 10.3847/2515-5172/ad148d

Werner, S. V., Hatch, N. A., Muzzin, A., et al. 2021, Monthly Notices of the Royal Astronomical Society, 510, 674–686, doi: 10.1093/mnras/stab3484

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, The Astronomical Journal, 120, 1579–1587, doi: 10.1086/301513

Zabludoff, A. I., & Mulchaey, J. S. 1998, The Astrophysical Journal, 496, 39, doi: 10.1086/305355

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

Zhao, D., Aragón-Salamanca, A., & Conselice, C. J. 2015, Monthly Notices of the Royal Astronomical Society, 453, 4445–4456, doi: 10.1093/mnras/stv1940

Zinger, E., Dekel, A., Kravtsov, A. V., & Nagai, D. 2018, Monthly Notices of the Royal Astronomical Society, 475, 3654, doi: 10.1093/mnras/stx3329

Appendices

Appendix A

Paper 1 appendix

A.1 Optimizing the WWFCS field positions

As mentioned in Section 3.3.1, the WWFCS performs observations by arranging 2 degree fields into a mosaic pattern, covering the cluster core, infall region and outskirts (Figure 3.1). To optimize the observational strategy, we aim to design the field positions in a way that maximizes the cluster coverage. Firstly, we place two fields at the core of each cluster, the region of the highest number density. This is so we can maximise 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

$$\frac{\text{Area within } 5R_{200} \text{ for N - 1 fields}}{\text{Area within } 5R_{200} \text{ for N fields}} > 97\%, \tag{A.1}$$

then we can remove one field, (we use 97% 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% threshold is exceeded. Starting from a 'naive' geometric tiling pattern (illustrated in Figure A.1, 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 optimised tiling method (illustrated in Figure A.1, bottom) the total number of fiber hours is reduced to 130 390. For the example shown in Figure A.1, even though we have removed four fields, we are still covering out to $5R_{200}$.

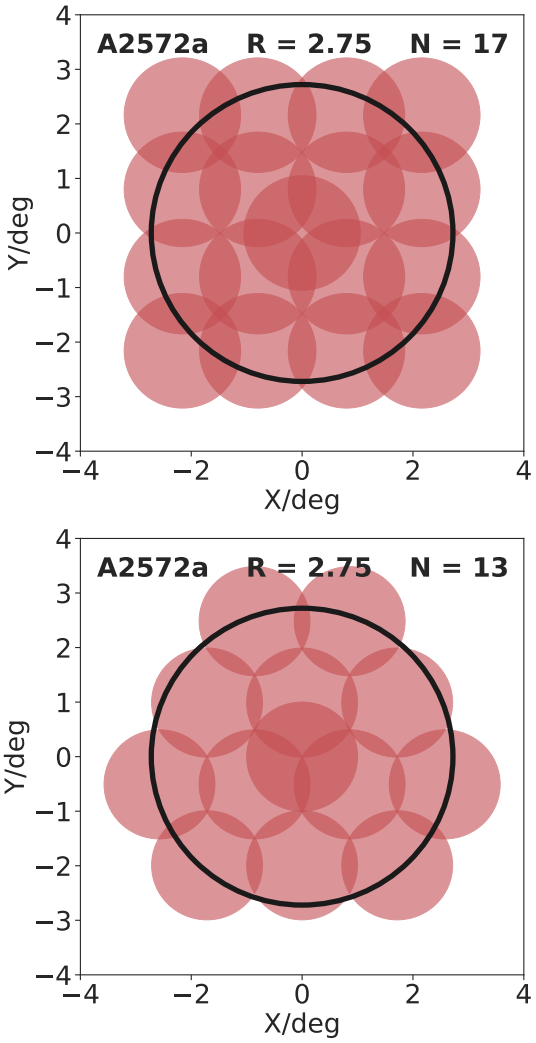


Figure A.1: An example field layout of cluster A2572a before this work (top) and after this work (bottom). The red circles are individual 2-degree diameter WEAVE fields whilst the black outer circle represents the angular diameter corresponding to $5R_{200}$ of this cluster. The numbers displayed are the total number of fields required to cover this cluster N, the cluster redshift and R_{200} taken from (Moretti, A. et al., 2017).

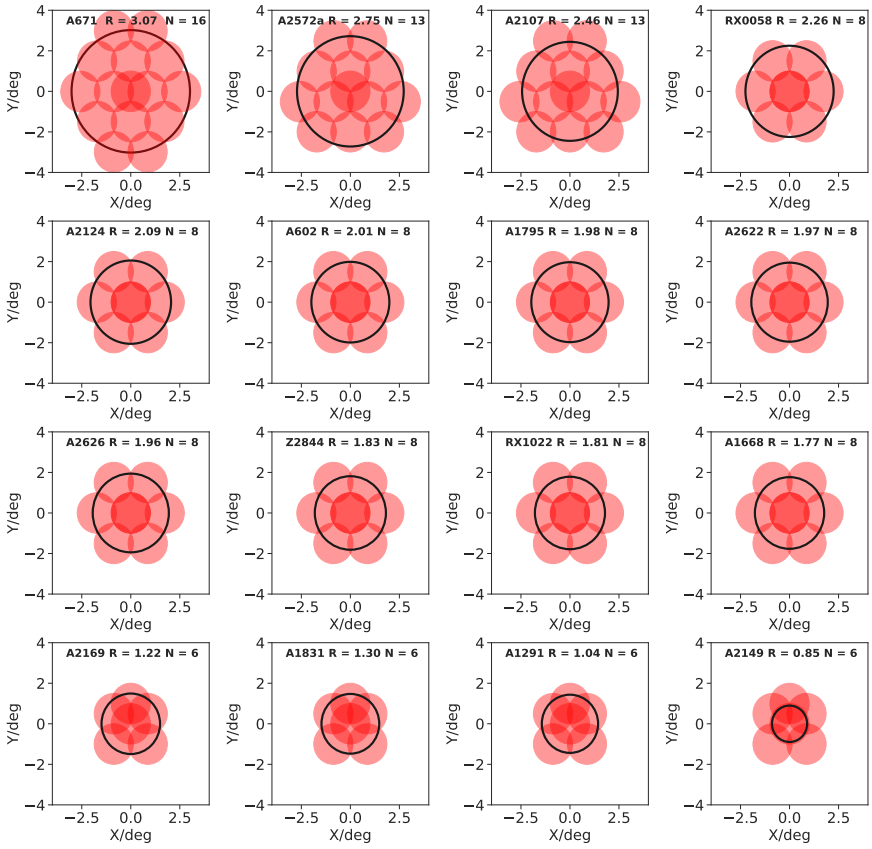


Figure A.2: WEAVE field pattern for each WWFCS cluster. The caption in each panel states the name of the cluster, $5R_{200}$ in degrees and the number of fields N used for each observation.

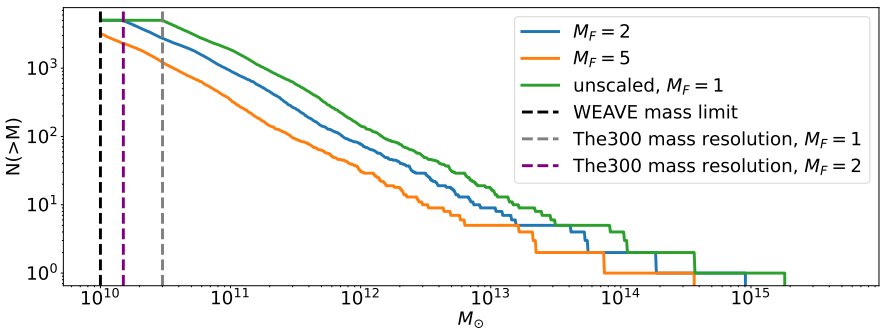


Figure A.3: Cumulative sum of individual halo masses of simulated galaxies for one galaxy cluster in `TheThreeHundred` that has been modified using the three M_F. All haloes follow the criterion set for selecting 'high–quality data' from the simulations, as described in Section 3.3.3. Different colours represent different scale factors that have been applied to the simulated catalogues. The vertical dotted lines represent different mass resolutions.

Of the clusters, 12 out of 16 have full coverage out to $5R_{200}$, whilst overall we have lost a total of 0.06% area within $5R_{200}$.

A.2 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 3.3.3.1.

Firstly, 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 Figure 3.2, 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 halos for each cluster. Figure A.3 displays the cumulative number of simulated halos 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 don't scale the simulations, ($M_F = 1$), the mass resolution limit is that of the dark matter particles in the simulations (Kuchner et al., 2021), given by the dotted grey line in Figure 3.4. For $M_F > 1$, we reduce the simulation mass resolution by dividing it by the mass scaling factor M_F. Whilst 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 observational limit $M_{\rm obs} = 10 \times 10^{10} M_{\odot}$, (corresponding to the r band limit used for WEAVE: $r_{\rm total} < 19.75$ which is equivalent to a stellar mass of $10^9 M_{\odot}$, (Kuchner et al., in prep)). At $M_F > 3$, we artificially lose halos 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 Figure 3.2, 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 for example Poggianti et al. 2010).

A.3 DisPerSE input parameters

As mentioned in Section 3.3.4, to extract the filament networks with DisPerSE we need to set a persistence threshold. We introduce a metric Ψ with 3 key parameters:

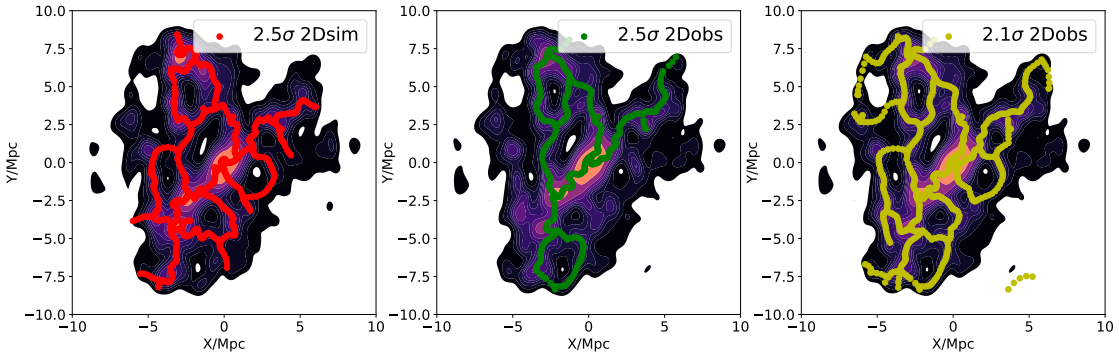


Figure A.4: This figure illustrates how the change in the density field when 'configuring' a cluster makes a change in the persistence parameter necessary when extracting the filament network. Left panel: KDE-smoothed density field of a simulated cluster analogue to WWFCS target RX0058, with the filaments 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 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.

1) $D_{\rm skel}$ median (the positional difference in the reference network spine and the configured network spine),
2) $D_{\rm skel}$ ratio: the ratio of the two methods of calculating $D_{\rm skel}$,
3) 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, whilst 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 not only map out 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 Figure A.4. 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.

Appendix B

Paper 2 appendix

B.1 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 Figure B.1. 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 4.5.3.

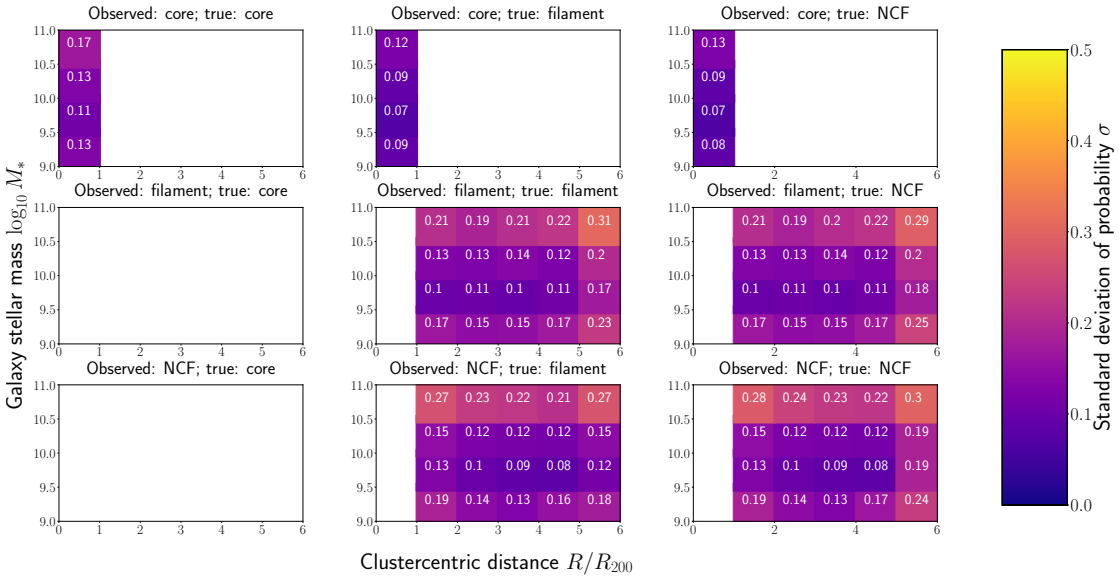


Figure B.1: The cluster-to-cluster scatter (standard deviation) of the probabilities shown in Figure 4.4.

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 B.1: 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 Figure 4.4. For each mass and distance bin, three numbers are given, each corresponding to each one of the columns in Figure 4.4. A number '−1' indicates that a given situation is not possible, and corresponds to a white cell in Figure 4.4. 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 Figure 4.4. We provide them here in tabular form for easy access.

Core	$0 < r < 1R_{200}$	$1R_{200} < r < 2R_{200}$	$2R_{200} < r < 3R_{200}$	$3R_{200} < r < 4R_{200}$	$4R_{200} < r < 5R_{200}$	$5R_{200} < r < 6R_{200}$
$10^{9.0} < M < 10^{9.5}$	0.69/0.14/0.18	−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

Table B.2: Probability of galaxies being identified in different cosmic web environments for high mass model clusters ($5.0 \times 10^{14} M_{\odot} < M_{\rm cluster} < 1.2 \times 10^{15} M_{\odot}$). The format is the same as in Table B.1.

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 B.3: Probability of galaxies being identified in different cosmic web environments for intermediate mass model clusters ($2.5 \times 10^{14} M_{\odot} < M_{\rm cluster} < 5.0 \times 10^{14} M_{\odot}$). The format is the same as in Table B.1.

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 B.4: Probability of galaxies being identified in different cosmic web environments for low mass model clusters ($2.5 \times 10^{14} M_{\odot} < M_{\rm cluster} < 5.0 \times 10^{14} M_{\odot}$). The format is the same as in Table B.1.

Appendix C

Paper 3 appendix

C.1 Non mass-weighted networks

For comparison, we repeat the process of determining the distance between each cosmic web node to its nearest galaxy group but without mass-weighting. Overall, the matching is much less successful than in the mass-weighted case: without mass-weighting we only find 43% of the cosmic web nodes match galaxy groups. We demonstrate this in Figure C.1. The presence of a third peak, seen in the lower left histogram at distances $R/R_{200} \sim 10^{-1}$, is due to the node latching on to a subhalo within the group halo. This result implies using mass-weighting in the filament finding process very significantly improves our ability to locate galaxy groups using network nodes.

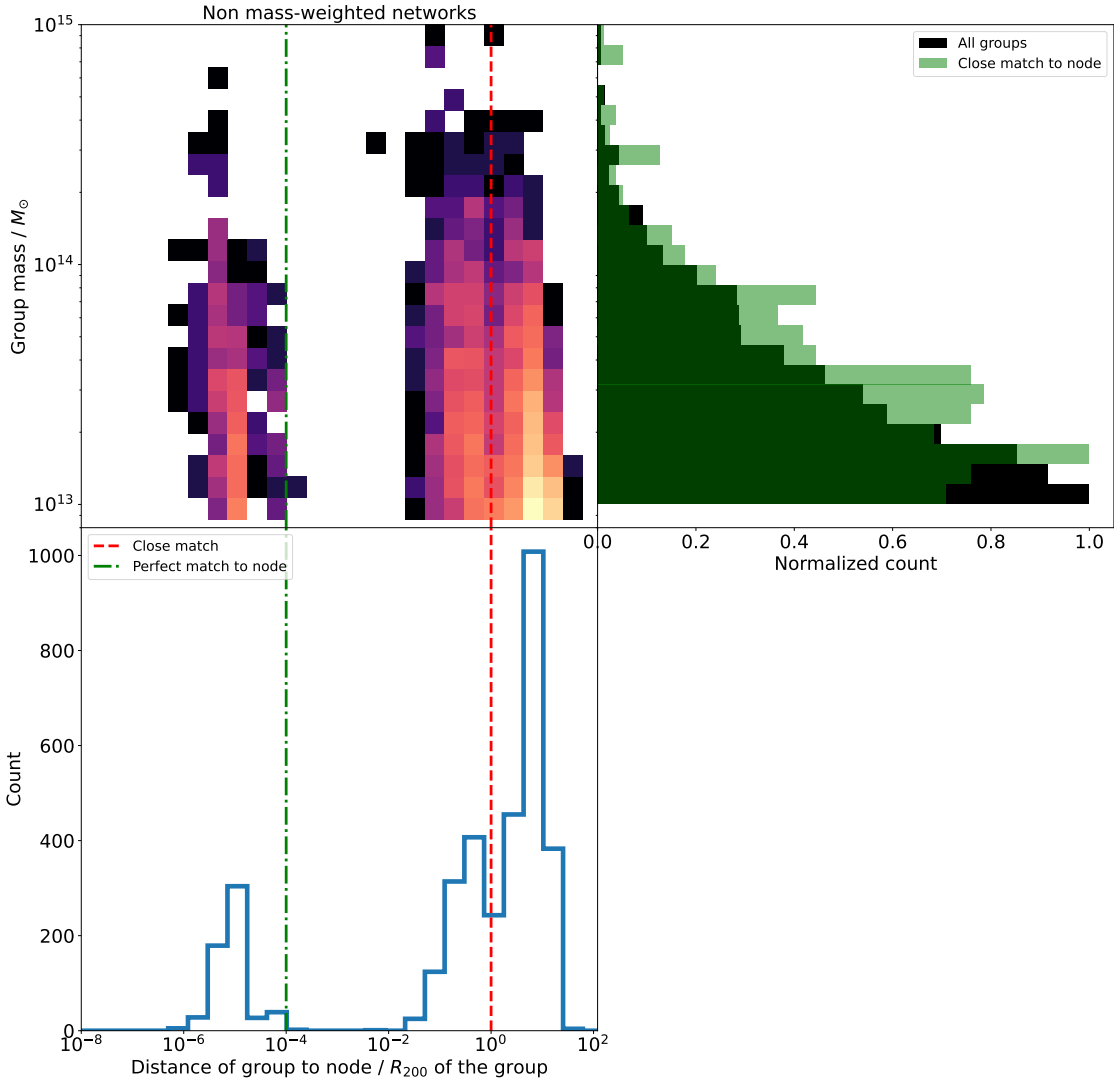


Figure C.1: The same plot as Figure 5.3 but for the non mass-weighted case, completed in 3D. In the right panel we show the mass distributions of the entire group sample and compare it to the sample of groups that are close matches to nodes, as done in Figure 5.6.