Rijksuniversiteit Groningen

Bachelor Research Project

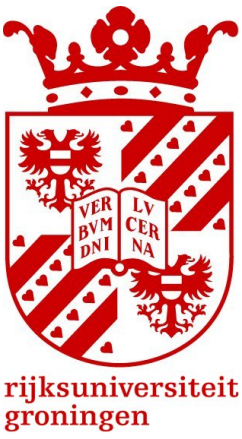
COSMIC WEB CLASSIFICATION: NEXUS+ VS. DISPERSE

How does DisPerSE, a topological based algorithm differ to the scale-space, geometrical approach taken by NEXUS+ upon analysis of the cosmic web?

Author:
Erwan Hochart *(s3424596)*

Supervisor:
prof. dr. Rien van de Weijgaert

July 3, 2020



Abstract

By describing the matter distribution of the Universe at the largest of scales, ingrained within the cosmic web is an abundance of information on the history of structure formation within our Universe. However, due to its non-linear nature and ambiguous definitions in defining the different constituents, cosmologists have difficulty analysing the cosmic web, leaving essential information out of reach. Over the decade's numerous formalisms have been developed to try and extract this hidden information. The report focuses on two formalisms, namely DisPerSE and NEXUS+, and begins with dissecting the mathematical formalisms they incorporate in their algorithm. More explicitly, DisPerSE uses a topological approach in its analysis of the cosmic web while NEXUS+ adopts a scale-space approach. After building up the required knowledge to get a grasp on the algorithm and how they differ, simulations were conducted on a pre-defined Universe to compare both qualitative and quantitative results. The results found that although DisPerSE is more mathematically robust, it has issues when tracing the different environments due to its topological approach being unable to preserve the shape and size of the various constituents. Contrariwise, results showed that NEXUS+ was successful in the delineation of the environments and rendered better results due to its geometrical approach in identifying the different regions resulting in better quantitative results as well. The success NEXUS+ has in portraying the different constituents combined with the scale-space approach taken to probe more efficiently the intricacies within the cosmic web suggests that NEXUS+ provides a better analysis of the cosmic web with its scale-space, geometrical approach.

Acknowledgements

To begin, I would like to thank my supervisor Rien van de Weygaert who, although was extremely busy given the circumstances, still found time to discuss and give ideas as to how I should proceed in the report in a manner which could only motivate one to do more.

I would also like to thank Roi Kugel for his in-depth explanation towards certain concepts as well as discussing certain results with me to bounce ideas off one another.

To Leon, Martin and Eite, thank you for being extremely patient with me and my numerous problems upon using and compiling DisPerSE. Your help was extremely helpful and I am grateful for it especially since at first glance using such a software seemed an impossible task given all the problems.

Finally I'd like to thank both Robin Kooistra and my dad for helping me with the programming side of the project and explaining how to actually use DisPerSE as well as helping me create more efficient code and how to use the command line to probe processes as well as extract more information.

Contents

1 Introduction

The cosmic web describes the distribution of matter on scales of tens to hundreds of mega-parsecs induced as a result of the minute perturbations in the primordial Gaussian density field of the Universe as well as gravitational instability (Bond et al., 1996). Upon observation, four key characteristics that define the cosmic web have emerged raising complexities for a tool to rigorously analyse the environment.

The cosmic web consists of rich anisotropic structures all interconnected with one another moulding the wispy-like pattern observed today. For instance vast, empty voids dominate the volume content of the Universe and engulf the sheet-like walls, whom themselves envelope filamentary networks that branch off from clusters and act as a highway that distributes the matter content of the Universe. These morphological elements together constitute the intricate web-like pattern observed today, and though it may seem stochastic, countless N-body simulations have shown that this pattern arises naturally through gravitational instabilities (Gott (1987); van de Weygaert and Bond (2008)). Another observation is that the cosmic web is a multi-scale environment and this results from the hierarchical nature in structure formation, in which small structures form and over time coagulate with one another merging into more massive structures as illustrated in figure 1-1 giving for a range of possible densities the different environments can comprise of. As the Universe evolves, gravity causes these environments to clump towards denser regions and a vast asymmetry between overdense and underdense regions emerges, giving the cosmic web another defining characteristic since the lesser dense regions end up dominating the volume composition of the Universe.

To date, the Sloan Great Wall is the largest structure observed measuring at around $400h^{-1}$ Mpc (Gott et al., 2005). On the other hand, the oldest example of a morphological element was detected as a filament by the Subaru survey at a redshift $z \approx 3$ (Matsuda et al., 2005) providing strong evidence of the existence of these pronounce cosmic structures at earlier epochs. These multi-scale observations of the cosmic web, as well as the knowledge of it being a prominent feature at earlier aeons, has instigated an abundance of research to be conducted due to its link with cosmic structure formation processes.

Since the pattern observed is induced by the primordial Gaussian density field, analysis of the morphological features emerging from the cosmic web allows for information on the nature of dark matter and dark energy as well as the formation and evolution history of galaxies to be obtained. For instance, voids have been used as a testing site to probe the nature of dark energy and dark matter as well as test theories of gravity (Lavaux and Wandelt (2010a)), while filaments may hold essential information on the missing baryonic matter component of the Universe (Fukugita et al., 1998). To extract the plethora of information ingrained within the cosmic web a multitude of studies have been conducted over the decades using both N-body simulations as well as galactic redshift surveys.

By analysing DisPerSE, a topologically motivated formalism and that of a geometrical one

provided by NEXUS+ which incorporates a scale-space analysis, it's found that NEXUS+ has more success in delineating and identifying the different environments, especially the tenuous structures and subsequently provides more reliable qualitative and quantitative information on the cosmic web as opposed to DisPerSE.

1.1 Outline of the paper

This thesis is composed of five main sections, in section 2 the constituents of the cosmic web as well as some defining features of the cosmic web are introduced to provide the reader justification as to why the cosmic web is so hard to study. Section 3 will dissect the mathematical theories behind both formalisms providing a better understanding as to why the discrepancies in their results emerge. Afterwards, a description of a chosen model Universe in section 4 will be given followed up by a comparison between quantitative and qualitative results obtained using both formalisms in section 5. The results obtained will be used to support the final conclusion of the paper given in section 6, which will also mention possible room for improvements in the study.

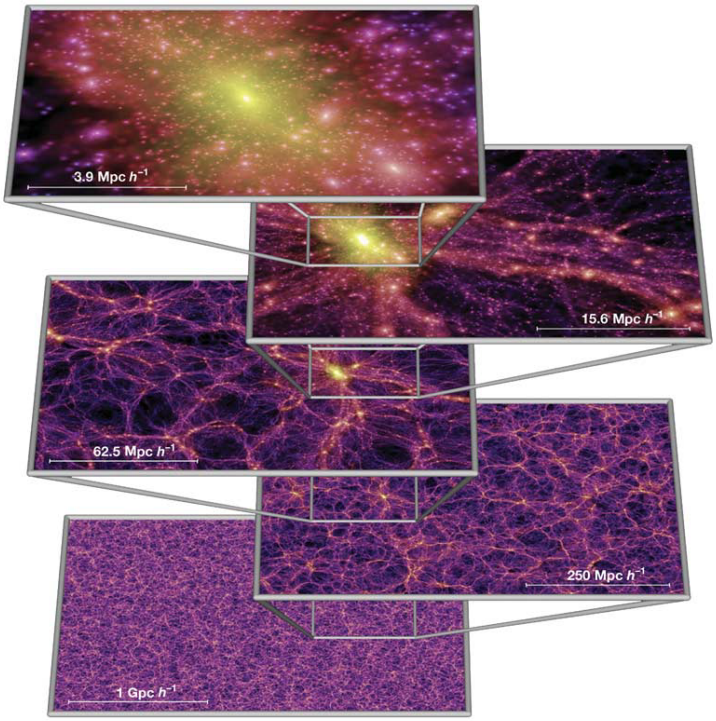


Figure 1-1: Structures not yet fully developed are composed of denser smaller sized substructures formed at an earlier time making it a prime example of the hierarchical development of the structure. The image illustrates this by showing clusters being composed of galaxies and dwarf galaxies, while at larger scales they can form super-clusters situated at nodes, or if the distribution is sparse enough, filaments or walls and producing the intricate web-like pattern observed at the largest of scales. (image courtesy of Springel (2005))

2 The Cosmic Web

The Sloan Digital Sky Survey (SDSS) (Fukugita et al. (1998); Gunn et al. (1998)) shows that at the mega-parsec scale, the spatial distribution of galaxies and clusters of galaxies form a wispy-like environment composed of different constituents. This intricate pattern emerges due to the primordial density field of the Universe and gravitational instabilities.

Numerous papers have shown that the Universe had a primordial Gaussian random density field expressed by a power spectrum of density fluctuations (Adler (1981); Bardeen et al. (1986); Bennett et al. (2003)) which also manifests itself in the observed temperature fluctuations of the Cosmic Microwave Background (Planck Collaboration et al., 2014). The second mechanism behind the web-like pattern observed is due to the evolution of perturbations as a consequence of gravity (Peebles, 1980). Dense regions such as clusters or nodes become denser over time as they accrete mass and continue to contract onto themselves, while sparse environments such as voids expand and have their mass outsourced to denser environments. Based on the initial conditions induced from the primordial density field and how perturbations evolve, four key characteristics describing the cosmic web have emerged upon its observation, making it hard for algorithms to analyse it rigorously.

First and foremost, the different morphological environments form complicated geometrical patterns giving its rich anisotropic nature, meaning identification of environments cannot stem solely from the appearance. These components not only have an anisotropic nature but a hierarchical one too leading to it's multi-scale nature in terms of mass and size. The idea is that structures emerge out of less massive but denser substructures which end up coagulating with one another giving for no well-defined structures at a particular scale or density, in turn causing the delineation of the different environments extremely difficult as there is no optimal scale at which to analyse the cosmic web as well as the fact that environments comprise of a range of different sizes and densities. In fact, upon observations the cosmic web spans six orders of magnitude in densities, with the density distribution of different environments tending to overlap with one another as illustrated in figure 2-1. This overlap of their density distribution means that an algorithm can't identify environments purely based on their densities either. These environments are all interconnected with one

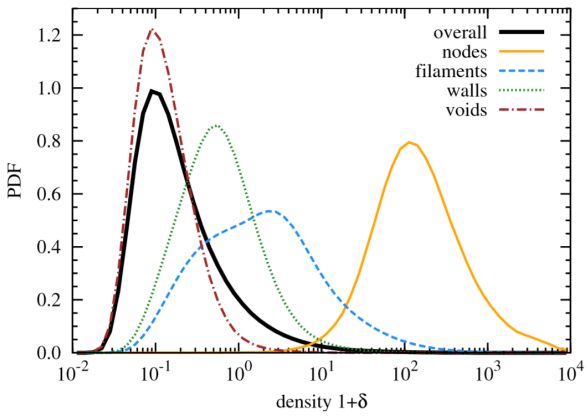


Figure 2-1: Probability density function (PDF) graph of the different morphological environments in the cosmic web as found by Cautun et al. (2014) using NEXUS+.

another forming the last key characteristic of the cosmic web since clusters of galaxies end up forming the nodes of the environment in which filaments branch out from, diverging into sparser void regions. These filamentary networks are enveloped by the sheet-like walls who in turn get engulfed by the surrounding void regions giving for a high connectivity between the different environments causing the environment to be intricate by nature.

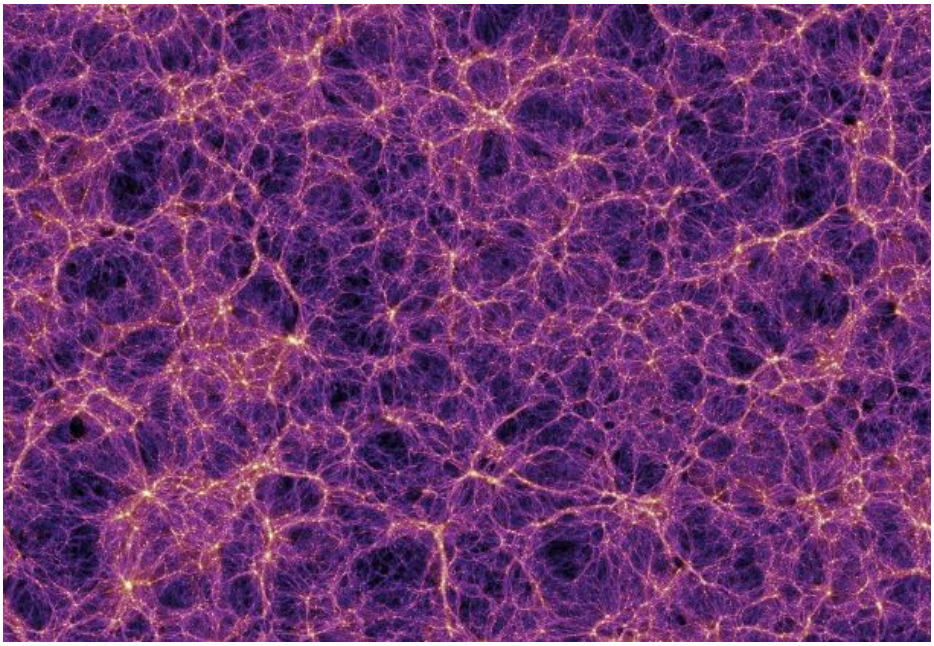


Figure 2-2: The figure shows the morphological components making up the cosmic web. Bright yellow spots show the denser regions and represent clusters. Filaments connect clusters situated at the nodes and form a network, all the while vast empty regions called voids dominate the volume of the Universe and engulf them. (image courtesy of the Millennium Simulation Project).

To better understand the different constituents of the cosmic web, a few terminologies are introduced. Figures 2-2 and 2-3 are also added to support these definitions.

Definition 2.1 (Voids): Voids are underdense regions and practically devoid of galaxies. They dominate the volume fraction of the Universe (Cautun et al., 2014) and come in a range of sizes, anywhere between 20-50h^{-1}Mpc (Libeskind et al., 2018). Our galaxy neighbours such a void called, rightfully so, the Local Void, whose diameter is 30h^{-1}Mpc (Tully, 1987). Their spherical shape and structure are suggested to provide a direct reflection on the nature of dark energy (Lavaux and Wandelt (2010b); Lavaux and Wandelt (2012)) providing a natural environment to investigate such phenomena, and as alluded to in section 1, also happens to be a good testing ground for gravitational theories due to its lack of matter. These regions engulf the elongated filaments, sheet-like walls and dense clusters of the cosmic web.

Definition 2.2 (Walls): Walls are the most tenuous coherent structure observed in the large-scale Universe, making them difficult to detect even with the help of N-body simulations. These structures have undergone gravitational collapse only about one of its

axes, causing its sheet-like appearance and form the boundary between filaments and voids. They contain low luminosity galaxies and have low densities making them hard to observe as well as less prominent a feature compared to filaments (Cautun et al., 2014).

Definition 2.3 (Filaments): Filaments are structures that have undergone gravitational collapse along two axes and are the most prominent feature of the cosmic web. They are the elongated structures that branch off one another and tend to distribute matter from underdense regions to overdense ones, acting as the cosmic highway for matter.

Definition 2.4 (Clusters): Though stars and galaxies make up some portion of their mass, clusters are often called dark matter haloes due to dark matter dominating its mass composition. These overdense structures form the nodes of the cosmic web environment in which rich clusters can have up to several thousands of galaxies in a relatively small volume of only a few mega-parsecs (van de Weygaert et al., 2009).

Though easy to distinguish by eye, these different structures are difficult to analyse computationally due to the characteristics of the cosmic web leading to a lack of mathematical and clear definitions which to describe them as. Though clusters can be defined mathematically by being gravitationally stable environments, filaments and walls have yet to have a rigorous definition attributed to them. Nonetheless, an abundance of research is conducted to attempt to obtain the copious amount of information ingrained within the cosmic web with the use of both observational data and of N-body simulations, a tool necessary in the investigation at the non-linear regime, induced by the ever-growing density perturbations of the cosmic web.

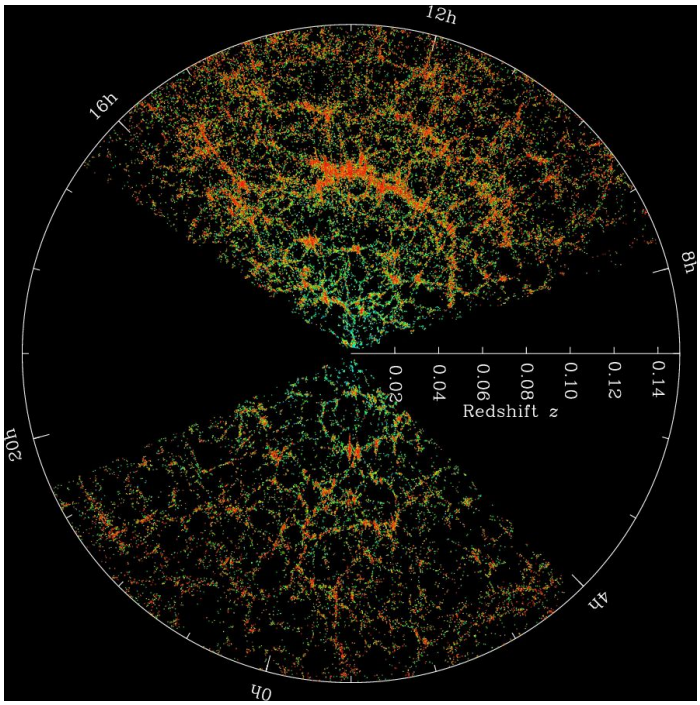


Figure 2-3: The Sloan Digital Sky Survey showing the distribution of galaxies up to a redshift of $z \approx 0.15$. Filament networks, as well as vast empty voids and cluster regions, are observed and help illustrate the connectivity between the environments as well as the pattern that makes up the cosmic web. (image courtesy of Fukugita et al. (1998))

3 Theory

3.1 Methods of Analysis for the Cosmic Web

At first order, the Universe is composed primarily of dark energy and dark matter due to its baryonic mass fraction being insignificant. This approximation allows for more efficient computation since there is a vast amount of different baryonic processes such as stellar formation that no longer need to be processed. The other two mass constituents, dark energy and dark matter are only linked to one process be it the cosmic expansion rate of the Universe or gravitational interactions. This simplification leads to two different types of N-body simulators; a dark matter simulation and a hydrodynamical simulation (Cautun, 2014). This report focuses on the first class of simulations, which has the advantage of being processed quickly as well as being simple to implement as it only takes into account the gravitational interactions between halo particles.

To extract the information ingrained within the cosmic web a variety of algorithms have been developed over the decades. González and Padilla (2010) developed an algorithm that analyses the morphological environments based on particular physical criteria with the help of a tessellation-based density estimator. Genovese et al. (2010), on the other hand, used a non-parametric formalism utilising the medial axis of the data distribution to represent the filamentary network. Nowadays, most modern algorithms attempt to implement the morphological information rooted in the gradient of the Hessian density or potential field. The two formalisms the report investigates makes use of the density field, allowing for a more thorough analysis of the multi-scale nature of the matter distribution at the cost of noise sensitivity as well as the limitation of not being directly correlated to the underlying dynamics of structure formation, which the tidal field otherwise is (Cautun et al., 2013).

Since it is hard to provide a conclusive comparison on two formalisms that analyses an environment so ambiguous in form, the following points give a general overview on what makes an algorithm successful when tracing the cosmic web:

- Successful in tracing the defining features and therefore rendering of the anisotropic pattern exhibited in the cosmic web for both a discrete particle distribution as well as observational data.
- The algorithm should be able to account for the hierarchical nature of the cosmic web to provide users with a user-free, adaptable fashion of identifying environments.
- The algorithm should be able to examine the multi-scale nature of the cosmic web by identifying both significant structures as well as the tenuous ones.

The first point is particularly important since many errors manifest themselves in observational data. Errors such as projection effects like Finger-of-God or magnitude limitations can all affect one's data. N-body simulations are therefore often used since they provide a fully sampled 6-dimensional phase space and density field for any given Universe at any given time, giving for a wealth of information for cosmologists who now

have free reigns on their data.

This paper investigates two Hessian density-based methods. DisPerSE[1], whose code is online for the public, looks at the topology of the field to delineate morphological environments while NEXUS+ takes a scale-space and geometrical approach in its analysis. The following subsections of the report will discuss the theory behind the two formalisms to understand how they analyse the cosmic web.

3.2 Delaunay Tessellation Field Estimator

Tessellations appear all around us, be it a brick wall or the gridded squares in your notebook, they happen to form a crucial part in both formalisms since by construct they are sensitive to the spatial distribution of the data points. This sensitivity to the spatial distribution of particles is useful for the cosmic web as it makes it sensitive to the anisotropic patterns displayed within the environment giving for more reliable morphological detections in a fashion that doesn't employ bias introduced by user parameters.

Delaunay Tessellation Field Estimator, or DTFE for short, is used by both NEXUS+ and DisPerSE and was developed by Schaap and van de Weygaert (2000) with the premise of constructing a continuous density field from a given discrete sample of N-body particles in a parameter-free manner. DTFE constructs the continuous density field by first processing the Delaunay and Voronoi tessellation of a given data set. From this, it then proceeds to linearly interpolate the density field based on the density values at the vertices of the contiguous Voronoi cell (figure 3-1).

To understand why it is so successful in doing so, a brief overview of Voronoi tessellations is given in Appendix A, while the following section will introduce the concept of tessellations with a focus on Delaunay tessellations. van de Weygaert and Schaap (2009) define a tessellation as:

Definition 3.2.1 (Tessellation): A tessellation in d-dimensional Euclidean space $\mathbb{R}^d$ is a set $\mathbb{T} = \{X_0, X_1 ... X_i\}$ of d-dimensional cells $X_i \subset \mathbb{M}^d$ such that:

1. $X_i \cap X_j = \varnothing$
2. $\bigcup_i X_i = \mathbb{R}^d$
3. $\#\{X_i \in \mathbb{T} : X_i \cap B \neq 0\} < \infty, \forall$ bounded $B \subset \mathbb{R}^d$

One can think of a tessellation as an arrangement of a finite amount of non-overlapping geometric structures, in this case, simplices (introduction in Appendix B), that span the complete d-dimensional Euclidean space $\mathbb{R}^d$ and can come in a variety of forms, for instance, Delaunay tessellations. A Delaunay tessellation is the partitioning of space into triangles whose vertices are formed by points within the data set. DTFE exploits a few properties

[1] http://www2.iap.fr/users/sousbie/web/html/indexd41d.html?

associated with Delaunay tessellations to construct its continuous density field (Aragón-Calvo et al., 2007):

1. Delaunay tessellations are sensitive to the local point density causing every cell to be unique.
2. Delaunay tessellations are sensitive to the local geometry of this sample point distribution.
3. A Delaunay tessellation will try and maximise its angles. It will try to maximise the smallest angle without necessarily decreasing its maximum angle.

These properties allow an adaptable reconstruction of a full volume covering and volume-weighted continuous density field. More precisely, the first property provides DTFE with an estimate of the local density based on the inverse contiguous volume of a specific cell (van de Weygaert and Schaap, 2000), while the second property allows DTFE to trace the anisotropic features exhibited in the cosmic web in a scale-free manner. Finally, the third property makes DTFE adaptive to a given sample set. This property makes DTFE a powerful tool for the analysis of the cosmic web as its adaptability provides sensitivity to all levels of the substructure present and therefore probes the hierarchical nature of the cosmic web.

The success of DTFE in reconstructing the density field is accentuated by its ability to not pollute the given distribution with any artificial features, which would otherwise introduce errors (Aragón-Calvo et al., 2010).

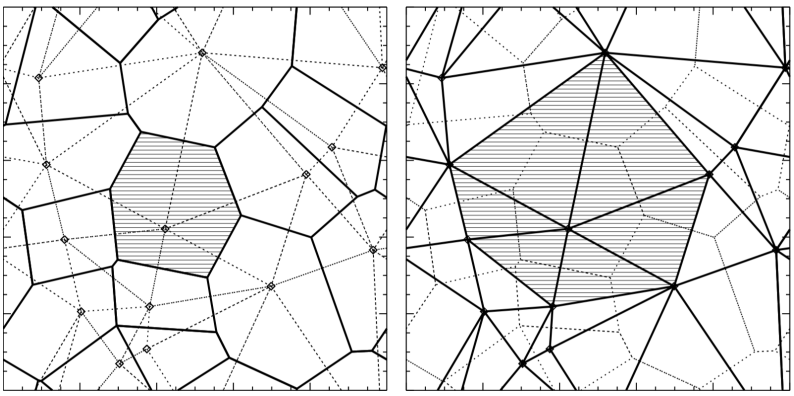


Figure 3-1: Left Frame: Illustration of Voronoi tessellations (solid lines) and Delaunay tessellations (dashed). Right Frame: The contiguous Voronoi cell is shaded in and shows all the Delaunay cells that overlap the Voronoi cell shaded in from the left hand image. (image courtesy of van de Weygaert and Schaap (2000))

By utilising properties of both Voronoi and Delaunay tessellations, DTFE has many advantages when compared to other reconstruction techniques. For one, DTFE reconstructs a volume-weighted and volume-covering density field spanning the volume of $\mathbb{R}^d$ in a way that preserves the local geometry of the point distribution. Some methods are

unable to do so and provide a grid rather than a continuous field. By reconstructing a continuous field, more information on the cosmic web can be extracted, providing researchers with better results. Another advantage DTFE has over other algorithms is that it implements volume-weighted averages rather than mass-weighted averages simplifying its computation a great deal (van de Weygaert and Schaap, 2000). Furthermore, other algorithms tend to be insensitive towards the sampling field causing far from optimal performance in both high and low-density regions. The user often fixes this issue in a non-natural manner which not only affects the computational resources needed but also introduces bias and errors within the data set (van de Weygaert and Schaap, 2000). DTFE bypasses this limitation due to exploitation of the minimum triangulation property making it adaptive by nature and further limiting the errors induced during the reconstruction of the density field in a way that preserves the hierarchical and multi-scale nature of the cosmic web.

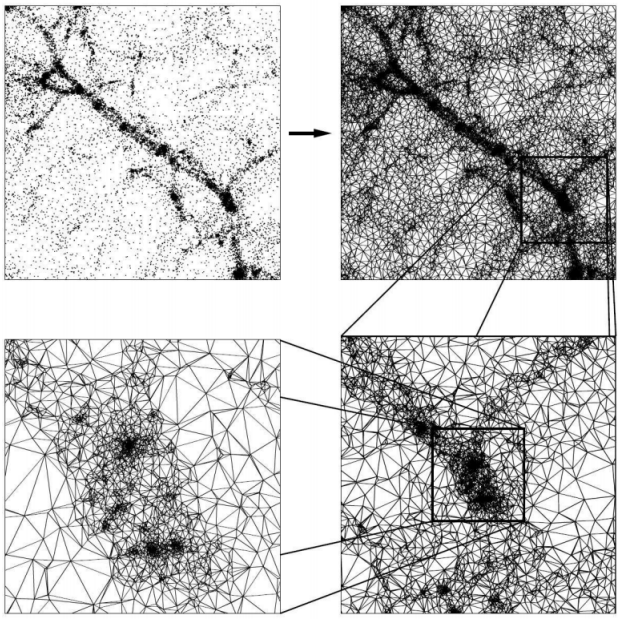


Figure 3-2: A sequence of images zooming in on a filament illustrating the adaptive nature of the DTFE. (image courtesy of van de Weygaert et al. (2009))

By being able to recover the density distribution for a wide range of magnitudes, DTFE has proven to be a useful computational tool upon analysis of the cosmic web given its multi-scale nature (van de Weygaert and Schaap, 2000). DTFE has also shown to successfully reproduce the anisotropic patterns observed in the cosmic web as well as take into account the hierarchical nature of the cosmic web without adding any artificial noise in a scale-free and adaptable fashion. All this amounts to a reliable reconstruction of the continuous density field and therefore makes it an essential algorithm for which both DisPerSE and NEXUS+ use when analysing the cosmic web.

3.3 Topological Approach

What is a hole? A hole is an abstract concept, one without precise meaning in the dictionary and only makes sense based on a given context. For instance, one understands that a cup has a hole which allows for liquid to flow in and out, however, if you continuously deform the glass and flatten it, at what point does this hole disappear? Does a bowl have a hole? Topology is a branch in mathematics that looks at fixing this ambiguity by relating holes with topological invariants, highlighting the underlying structure of a given manifold.

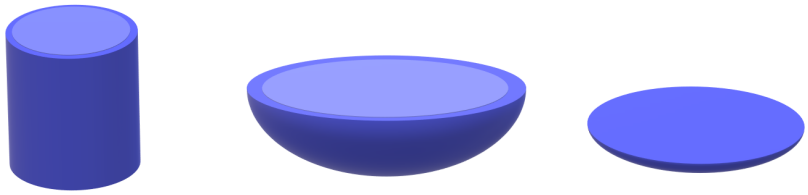


Figure 3-3: A sequence of images showing how a glass is topologically equivalent to a plate.

Topology studies the geometrical properties and spatial relations of surfaces which are invariant under continuous transformations. For the case of the mug and the doughnut in figure 3-4, one can continuously deform one of the objects to form the other, meaning they are topologically equivalent. The same reasoning follows as to why a mug and a glass aren't topologically equivalent, because both a mug and a glass have different types of 'holes'. For a mug to be deformed into a cup, a non-continuous transformation needs to occur either by adding or removing material.

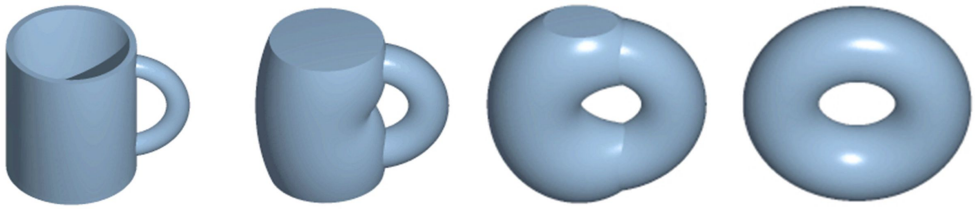


Figure 3-4: A sequence of images showing how a mug can be deformed continuously to form a donut and vice versa.

If two objects are topologically equivalent, as is the case between a glass and a plate as well as a mug and a donut, there exists a homeomorphism that describes the deformation of one surface to form the other. A homeomorphism is a bicontinuous and bijective function that maps elements on one manifold onto another. Formally this can be defined as (Gyulassy, 2008):

Definition 3.3.1 (Homeomorphism): A homeomorphism is a bijective and bicontinuous function that maps all elements in X onto Y, $f : X \to Y$. If such a function exists, then X and Y are said to be topologically equivalent.

The condition of bijectivity implies that there is a one-to-one correspondence for each point within the set and that no two points can share a point once mapped. Bijectivity, therefore, conserves the notion of continuity as it implies that material can neither be added nor subtracted in the transformation. By being bicontinuous, the function as well as its inverse are both continuous functions. Hence why it is possible to deform a mug into a doughnut and a donut back into a mug.

This branch of mathematics allows for quantitative analysis of the cosmic web since cosmologists can define environments and geometrical objects based on their invariants. Topological invariants such as the Betti number denoted as β_n (see Appendix D), allows for a quantitative description of the topology of a given manifold via the counting of k-cycles. Conceptually a d-dimensional Euclidean space will have d different types of holes, or Betti numbers, ranging in dimensions between 0 and $d-1$. For a 3D space the set of Betti numbers can be seen as (van de Weygaert et al., 2010):

- β_0 being the set of independent components in the manifold. These can be thought of as disconnected points.
- β_1 shows the number of independent tunnels. A tunnel can be thought of as a region of space in which going one direction and it's opposite will lead to no meeting of a boundary, but other directions will.
- β_2 denotes the number of holes or cavities within a surface. A shell, for instance, will have a β_2 number of one since it encloses a volume.

This extraction of topological information forms the foundation of DisPerSE since by analysing information on the topological properties of excursion sets, one can characterise the morphology of the different environments and delineate structures formed by the underlying density field with features induced by Poisson noise.

3.3.1 Morse Theory

Morse theory captures the relationship between the geometrical and topological properties of a function (Sousbie et al., 2008). The geometric properties of a function are ones that are unaffected by translations or rotations of the surface, for instance, how many edges a particular domain has or the location of its features. Topological properties, on the other hand, are invariant under homeomorphisms giving information on the connectivity of points. In cosmological terms, one may think of this as the number of galaxies in a given volume which is mathematically described by Betti numbers or the Euler characteristic and by analysing the evolution of topological invariants through different filtration levels allows for a delineation of structures formed by Poisson noise with ones formed by the underlying density field. DisPerSE also uses discrete Morse theory to partition a domain into a complex of ascending and descending manifolds to form a discrete Morse-Smale complex, providing a mathematical equivalence for each environment. To understand how and why this is allowed, a few definitions from Sousbie (2011) are given.

Definition 3.3.1 (kth Order Critical Point): A critical point is defined as a point whose derivative $\frac{df}{dx}$ is equal to zero. In d dimensions there will be $d+1$ different types of critical points. For instance, two dimensions has a minima (critical point of order $k=0$), a saddle ($k=1$) and a maxima ($k=2$).

Critical points show locations in which there is no preferential direction in the gradient flow. The order of the critical point is determined by the sign of the eigenvalues of the corresponding Hessian matrix $\mathcal{H}$.

Definition 3.3.2 (Morse Function): A Morse function is a smooth function f on a differentiable manifold $\mathbb{M}$ such that $f : \mathbb{M} \to \mathbb{R}$ while satisfying the following properties:

1. None of the critical points are degenerate. That is to say, for all x such that $\nabla f(x) = 0$, Det $\mathcal{H}_f(x) \neq 0$.
2. No two critical points of the function share the same value, making all critical points distinct from one another.

A Morse function is a function which links the critical points, the gradient and the topology of the level set. One can imagine this link between the three by thinking of a topographic map in which adjacent contour lines show how steep an ascent is and dots the peak of a mountain in a local neighbourhood. By removing a specific critical point or peak, one understands that the whole dynamic and shape of the topographic map changes. Morse theory, therefore, highlights the fact that all changes in the topology of a Morse function occurs around critical points on a given manifold (Gyulassy, 2008). This idea of investigating the evolution of level-sets also roots itself in persistence homology, a topological method used by DisPerSE to delineate structure formed by Poisson noise to those formed by the density field.

Though critical points have no preferential direction by definition, non-critical points within the function tell the gradient in which way to flow. Following the gradient flow leads to an important concept since it allows partitioning of space into the set of integral lines.

Definition 3.3.3 (Integral Lines): Integral lines, or field lines, are defined as a curve $L(t) \in \mathbb{R}^d$ where $\frac{dL(t)}{dt} = \nabla f(L(t))$ and satisfy the following conditions:

1. The integral lines origin and destination is a critical point. Mathematically this is denoted as $\text{org} f = \lim_{x \to -\infty} f(x)$ and $\text{dest} f = \lim_{x \to \infty} f(x)$.
2. No two integral lines can intersect. That is to say, no two integral lines can share the same point, however, they can share the same origin and destination.
3. The set of all integral lines spans $\mathbb{R}^d$.

The properties mentioned allow for a classification of every point in space based on their origin and destination. In turn, this allows for a partition in space based on the set of ascending and descending manifolds.

Definition 3.3.4 (Ascending/Descending d-manifolds): Given a Morse function defined over $\mathbb{R}^d$ with a critical point P of order k, the ascending $(d-k)$-manifold defines the region of space with dimensions $(d-k)$ in which the set of points reached by integral lines originate from P. The descending k-manifold shows the region of space with dimension k of the set of points reached by integral lines with destination P and is illustrated in figure 3-5.

The important concept to take from this is that an ascending or descending d-manifold always spans a domain of dimension d. In other words, a 0-manifold looks at vertices, 1-manifolds a line, 2-manifolds a surface and 3-manifold a volume (Sousbie, 2011). Subsequently, this provides a mathematical equivalence to the different environments since vertices are associated with clusters, lines to filaments, surfaces to walls and volumes to voids.

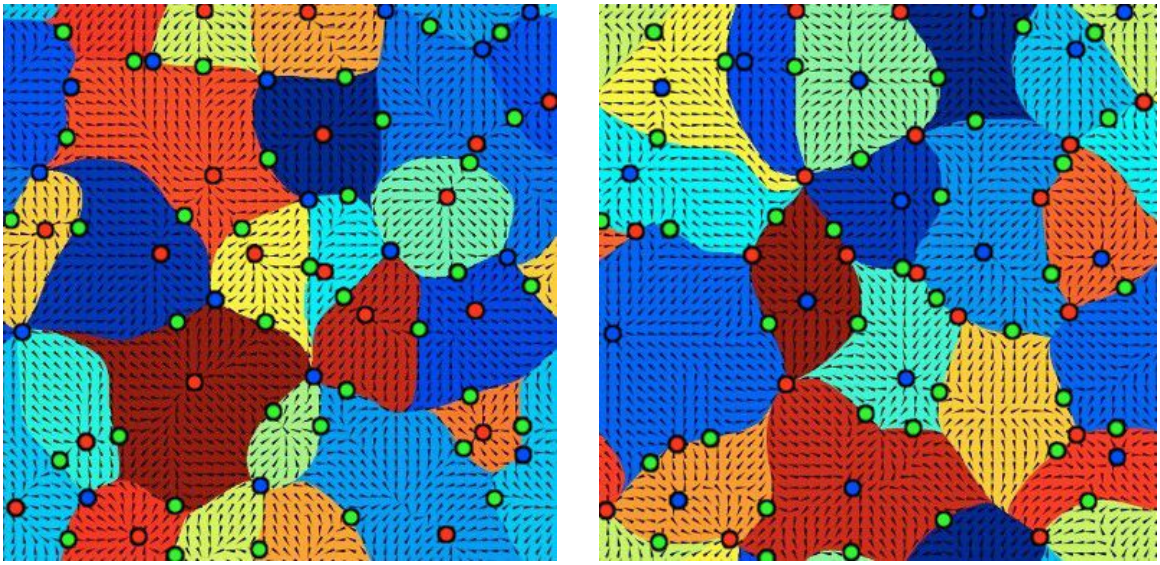


Figure 3-5: Left Frame: The set of ascending manifolds for a given complex. The red points denote critical points of order d (maxima), green points show critical points of order $(d-n)$ where $n \neq d$ signifying saddle points and blue points illustrating minima's, critical points of order 0. Right Frame: The set of descending manifolds for the same complex. Notice the link to tessellations in which the cells of the complex for either frame show the set of points closest to its nuclei (maxima or minima) based on the weighted gradient. (image courtesy of Sousbie (2011))

Figure 3-5 illustrates how Morse theory allows for partitioning of space into ascending and descending manifolds based on the gradient of the field. The set of its ascending and descending manifolds constitutes the Morse complex, which gets further dissected via the Morse-Smale function.

Definition 3.3.5 (Morse-Smale Function): A Morse function is a Morse-Smale function if its ascending and descending manifolds only intersect transversely.

This condition of transversal intersections ensures that the intersection between manifolds is defined at every location with the intersection between a p-ascending manifold with a

q-ascending manifold having a dimension of $n = \min(p, q)$. The set of points attributed to an intersection are called the Morse-Smale n-cells.

Definition 3.3.6 (Morse-Smale n-cell): A Morse-Smale n-cell is a non-zero intersection between a p and q-ascending manifold of a Morse-Smale function such that $n = \min(p, q)$. A 1-cell is called an arc, 2-cell a quad and a 3-cell a crystal.

For a 2-manifold, a 2-cell of the descending Morse complex corresponds to the maximum of the Morse function f, the 1-cell corresponds to the saddle points and the 0-cells the minimum. For an ascending 2-manifold the 2-cell corresponds to the minimum, the 1-cell the saddle points and the 0-cell the maximum.

Definition 3.3.7 (Morse-Smale Complex): A Morse-Smale complex is the set of all n-cells attributed to the Morse-Smale function f. Figure 3-6 illustrates the idea.

A Morse-Smale complex allows for a natural partitioning of space based on the gradient flow and is the set of integral lines with a common destination and origin. Another way to look at it is that a Morse-Smale complex is a complex made of both the ascending and descending manifolds defined by a Morse-Smale function and so each cell comprises of the set of integral lines with a common origin and destination.

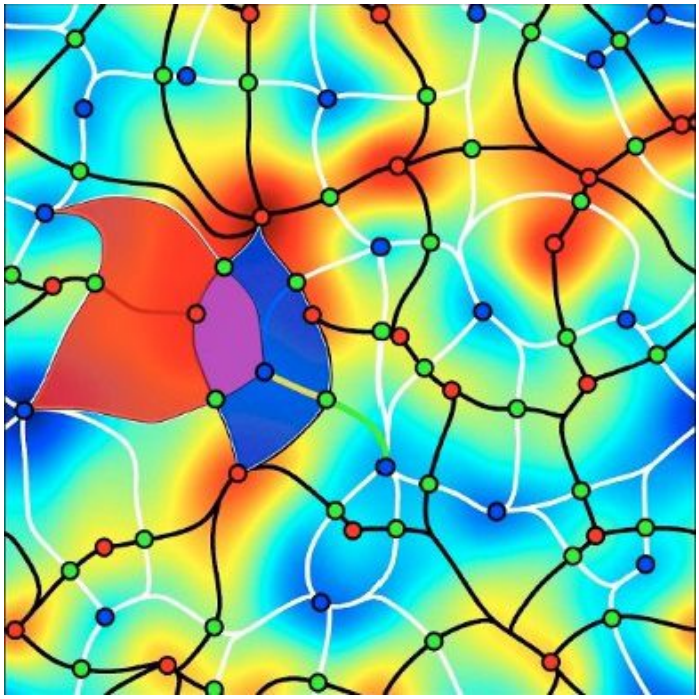


Figure 3-6: The Morse-Smale complex of the same Morse function that partitioned the space into a set of ascending and descending manifolds in figure 3-5. The purple region represents a quad and shows the intersection of a red descending 2-manifold with that of an ascending 2-manifold shaded in blue. The yellow segment in the blue region denotes an arc and represents the intersection of the green ascending 1-manifold with that of the blue descending 2-manifold. (image courtesy of Sousbie (2011))

3.3.2 Discrete Morse Theory

Though Morse theory gives for a mathematical equivalence between the environments within the cosmic web in a rigorous manner, observational data provides researchers with discrete data and so an algorithm needs to be able to partition space using non-continuous data.

Discrete Morse theory builds on the ideas implemented by Morse theory by using discrete functions defined over simplicial complexes which allow the construction of a continuous function by interpolating the field over a given cell. In turn, one can analyse the gradient field allowing for the partitioning of space into a Morse-Smale complex. This allows for a mathematically rigorous tool in which astrophysical data sets can be analysed from the distribution of different topological properties.

To understand how one transitions from continuous to discrete samples, a few definitions from Sousbie (2011), as well as an overview of simplices in Appendix B, are provided.

A discrete function f defined over a simplicial complex K associates a real-valued function $f(\alpha_k)$ for every simplex in the complex K.

Definition 3.3.8 (Discrete Morse Function): A discrete function is a discrete Morse function if and only if for each simplex $\alpha_k \in K$:

1. There exists at most one facet σ_{k-1} of α_k such that $f(\alpha_k) \leq f(\sigma_{k-1})$
2. There exists at most one co-facet β_{k+1} of α_k such that $f(\alpha_k) \geq f(\beta_{k+1})$

These properties ensure that the non-degeneracy of the Hessian matrix in Morse theory is conserved in discrete Morse functions such that in its local neighbourhood a simplex will have a higher value $f(\alpha_d)$ than its facets and a lower value than its co-facets. It gives a scalar value to every cell in the complex while the facets and co-facets of the cell determine whether it is critical or not.

Definition 3.3.9 (Critical k-simplex): A k-simplex α_k is critical for a given discrete Morse function provided the following conditions are met:

1. There exists no facets σ_{k-1} of α_k such that $f(\alpha_k) \leq f(\sigma_{k-1})$.
2. There exists no co-facets β_{k+1} of α_k such that $f(\beta_{k+1}) \leq f(\alpha_k)$.

Critical simplices act the same way as critical points. For two dimensions, a minimum is a critical vertex or 0-simplex, a saddle-point is a 1-simplex and a maximum a critical triangle, or a 2-simplex. By using the previous definition of a discrete Morse function, one sees that there are only two possible configurations for a simplex α_k. Either, one of its co-facets and all its facets have a lower value than a given simplex or all its co-facets and one of its facets have a larger value (Sousbie, 2011). This gives rise to a gradient flow as there is a defined preferential direction between a simplex and its co-facets (or facets) allowing for the partitioning of space into a discrete Morse-Smale complex.

The discrete gradient field that defines this complex obeys several properties (Sousbie, 2011):

1. If a simplex α_k has exactly one lower-valued co-facet β_{k+1} then $[\alpha_k, \beta_{k+1}]$ forms a gradient arrow.
2. If a simplex α_k has exactly one higher valued facet σ_{k-1} then $[\sigma_{k-1}, \alpha_k]$ forms a gradient arrow
3. The lowest valued $f(\alpha)$ simplex is the tail while the highest values the head, this entails that the gradient flows opposite to normal mathematical convention.
4. Critical simplices do not belong to an arrow and show the origin and destination of integral lines, making them analogous to critical points.

Condition three causes the previous definitions of ascending and descending manifolds (definition 3.3.4) to be reversed since it reverses the gradient flow. An ascending manifold now constitutes the set of integral who have a common destination while the descending manifold denotes the set of integral lines with a common origin giving for a final formal definition.

Definition 3.3.10 (Discrete Ascending/Descending d-manifolds:) Discrete ascending $(d-k)$-manifolds constitute the set of k-simplices that originate from a critical simplex α_k, or have a common minimum following the discrete gradient field. The discrete descending k-manifold, on the other hand, represents the set of k-simplices that have α_k as their destination.

A discrete Morse complex has the same definition as a continuous Morse-Smale complex (definition 3.3.7), the only difference being it uses the discrete ascending and descending manifolds. The computation of the discrete Morse complex is the end goal of the algorithm since it provides a natural delineation of the environment, thus making it a mathematically rigorous tool in the analysis of the cosmic web. The ascending manifold of a minimum (order $k = 0$) constitutes a three-dimensional cell, a 1-saddle point of a two-dimensional cell signifying a surface, a 2-saddle signifying an arc or line segment and a maximum ($k = 3$) which corresponds to a vertex, subsequently relating itself to voids, walls, filaments and clusters respectively.

Though discrete Morse theory gives for a mathematical equivalent for the various environments in a way that doesn't employ approximations, it comes with a few downsides. Computationally, the use of discrete Morse theory causes the processing of results to be time-consuming as well as using a large amount of computational resources. The topological method adopted is also unable to characterise the width of environments affecting its rendering of the different constituents when sampled on a grid if one wants to extract quantifiable information. Nonetheless, by implementing discrete Morse theory, DisPerSE has a reputation of delineating the underlying filamentary skeleton of a given data set with great success.

3.3.3 Persistence

Due to DTFE being both a scale-free method in the reconstruction of the density field as well as being unable to distinguish noise from features, DTFE has the limitation of exaggerating Poisson noise induced by the sample and therefore introduces artificial topological features (Sousbie, 2011). Dr Sousbie found that upon using DisPerSE without topological simplification, 94% of the identified features originated from noise (Sousbie, 2011). The abundance of false detections is due to discrete Morse theory capturing an excess of information about the underlying field, much of which can be deemed topologically insignificant. To extract only the most essential information as well as to remove artificial features in a meaningful way, DisPerSE uses the notion of persistent homology.

Persistence is a mathematical construct first introduced by Edelsbrunner et al. (2000) that delineates and removes noise from significant structures. Since critical points describe the structure of the level sets above or below a given filtration value, by analysing the evolution of the underlying field through different thresholds, one extracts the topologically significant structures.

For discrete functions, persistence is measured through filtration and analysis of the lifetime of different topological features, also called k-cycles. A k-cycle of a simplicial complex K (formally defined in Appendix D) is a k-dimensional topological feature in a d-dimensional space where $0 \leq k < d$. Analogous to the Betti numbers mentioned in section 3.3, in three-dimensional space a 0-cycle denotes a non-linked independent component, a 1-cycle a set of simplices that form a ring and a 2-cycle a shell that encloses some sort of volume. By seeing how these features evolve as one filters through the level set, quantifiable information on the structures of the cosmic web can be extracted. Formally, a filtration is defined as (Sousbie, 2011):

Definition 3.3.10 (Filtration): A filtration F of a given simplicial complex K is a sequence of $N+1$ sub-complexes K^i of K such that:

1. $\varnothing = K^0 \subseteq K^1 ... \subseteq K^N = K$
2. $K^{i+1} = K^i \cup \delta^i$

where δ_i is the subset of the simplices in K. In other words, only simplices with a value lower than the threshold τ_i, or $f(\alpha) < \tau_i$, belong to K^i and make up the sub-level set.

DisPerSE uses persistence by first applying a large threshold value onto the field. Taking function A in figure 3-7 as an example, by skimming the density field with $\tau_\rho = 22$ there is a single maximum detected. As the filtration threshold is decreased to $\tau_\rho = 21$, another feature is detected and remains unpaired until another feature is identified. The elder rule states that the most recently identified point be paired up with this recently detected topological feature, meaning that $\rho = 21$ and $\rho = 20$ are linked after the third filtration, as seen by the small figure between both the upper panel and lower panel graphs. The

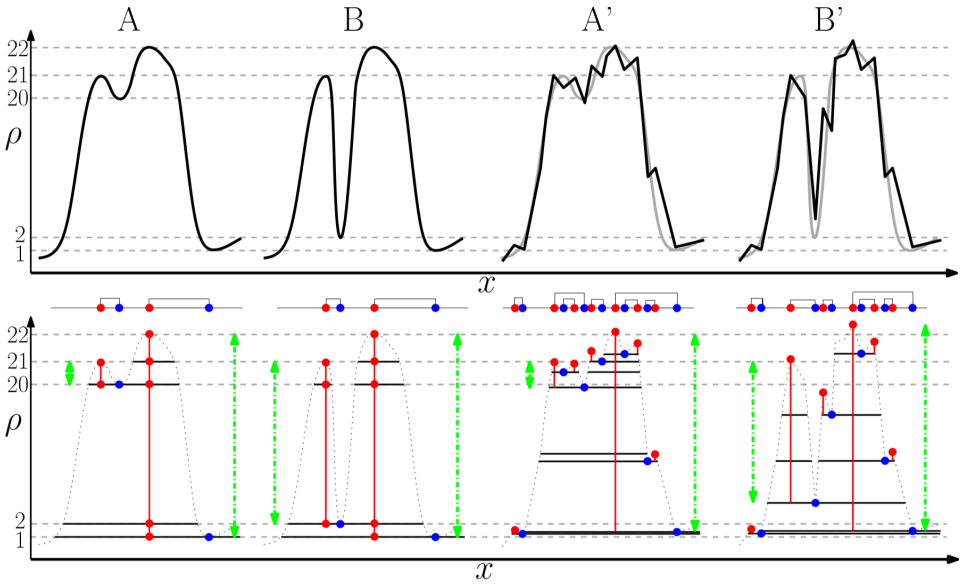


Figure 3-7: Two different data sets over a 1D function. The upper frame shows two functions (left) and a discrete distribution taken from these functions (right). The lower half of the figure shows the evolution of sub-level sets as the threshold $\tau(\rho)$ decreases. The green arrow emphasises the lifetime of certain topological features. The persistence pairs are denoted by the line squished between both frames. (image courtesy of Sousbie (2011))

persistence ratio shows the difference in ρ between the pair, so in this case, the persistence ratio amounts to 1. The larger the persistence ratio, the more significant a feature is, allowing the removal of noise in a mathematically robust manner.

As filtration of the field continues and the threshold decreases, the population of the sub-level set increases, with new k-simplices creating or destroying topological features. Newly detected simplices can fill in holes or create new ones, changing β_1 as well as having the potential to form volumes or destroy existing ones, in turn changing β_2. In other words, as the sub-level set grows, new features may be born and old ones may be destroyed. Persistence homology allows for filtration of noise as the k-cycle lifetimes attributed to noise are shorter than those of topologically significant structures (Zomorodian, 2005), which is reflected upon when comparing the discretely sampled graphs and the continuous ones in figure 3-7. More explicitly, notice that the links between pairs describing topological features for both graphs are larger for significant features when compared to the links between spurious ones.

3.3.4 Persistence in the Cosmic Web

Persistence allows a global delineation of the topological features with that of noise found in the cosmic web (Zomorodian and Carlsson, 2005). The interval length between a persistent pair allows DisPerSE to distinguish topological significant features with that of noise as it shows how much one has to change a function for a feature to be either removed or added (Gyulassy, 2008). DisPerSE implements persistence in its algorithm by providing the user

with a parameter, denoted by $n\sigma$, in which to filter the data. Any pair with a persistence ratio larger than $n\sigma$ will be conserved, while ones with a lower value eliminated from the data set. Figure 3-8 illustrates the strength of persistence in cosmological data sets by showing detected environments using persistence levels $n\sigma = 0, 2, 4$ and 6 for a given 256^3 particle Universe.

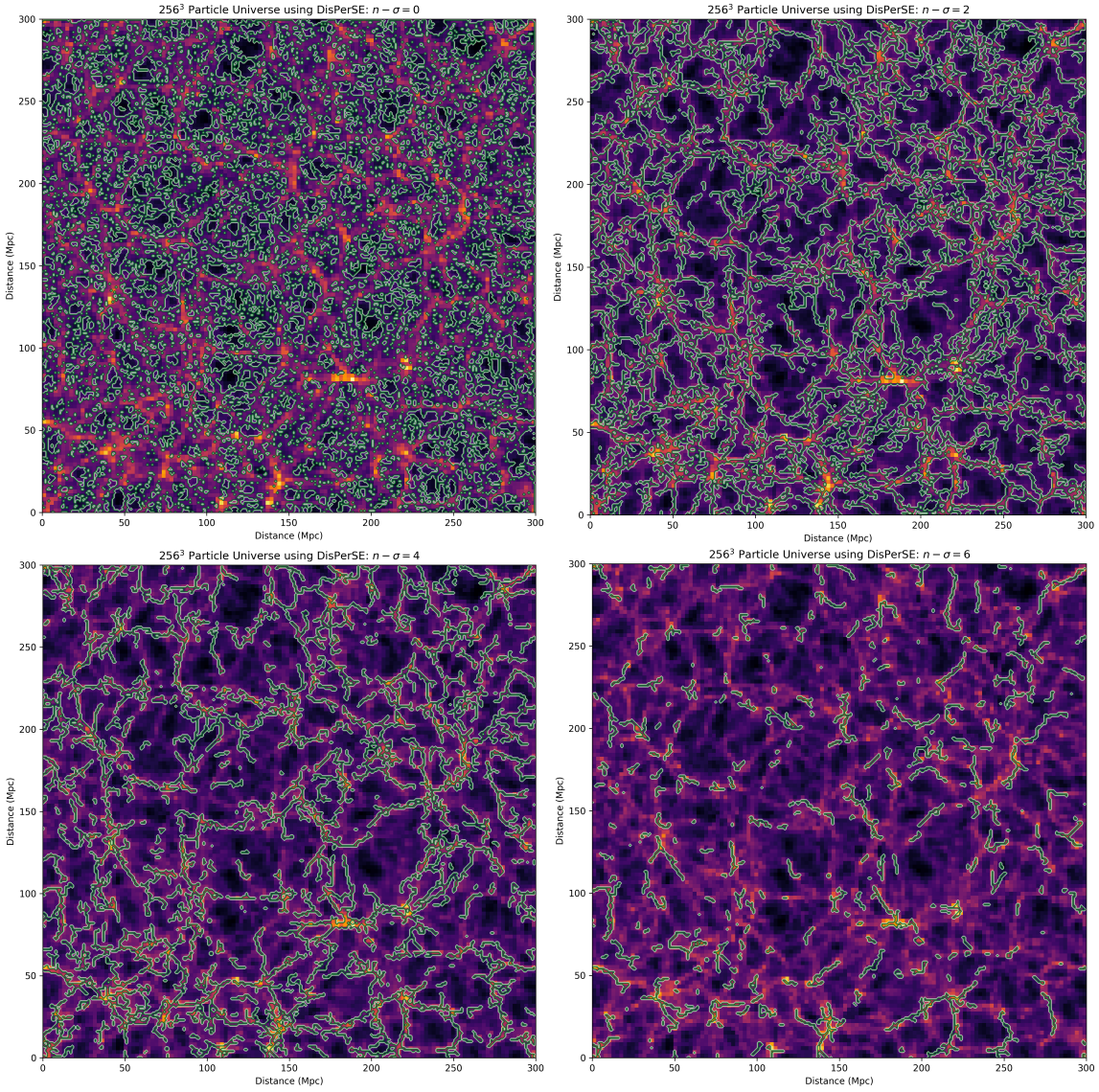


Figure 3-8: The detected filamentary network of a 256^3 particle Universe for several different $n - \sigma$ levels.

Figure 3-8 shows that for $n\sigma = 0$ the sample is covered with detected filaments and within two levels of $n\sigma$, the number of detections drastically reduces providing a coherent network

of filaments. This over-detection for $n\sigma = 0$ only reiterates the fact that via discrete Morse theory, most of the detected environments are caused by Poisson noise. In his original paper, Sousbie (2011) found that a $2-\sigma$ level has a 5% error false identification of significant features with that of noise while at $4-\sigma$ this value drops to 0.006%, showing the strength of persistent homology in the simplification of data sets.

Due to persistence analysing the lifetime of k-cycles, there are still spurious detections in voids, even at these relatively large $n\sigma$ as illustrated in the $n\sigma = 4$ frame with an identified structure at $(x, y) = (60, 190)$. This is understood since any form of noise in voids will have a large contrast in density with the background making the algorithm distinguish it as a significant feature. It also explains why errors occur more often in large voids where DTFE interpolates the density field over a large region of space since it causes Poisson noise to be more pronounced. Following the same reasoning, in regions where the density is high, significant structures may be labelled as topologically insignificant compared to their local neighbourhood and therefore eliminated from the sample as seen by the discontinuous filamentary network identified at higher persistence levels, as illustrated in figure 3-9. From this, one concludes that persistence trims topological features rather than only filtering out the noise, making it not adaptable when analysing the cosmic web.

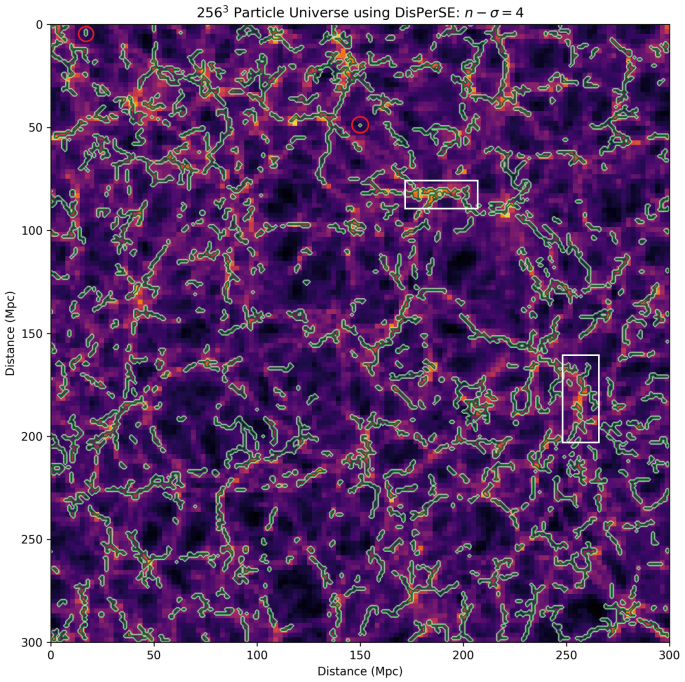


Figure 3-9: The detections marked in a red circle are spurious detections due to DTFE interpolating the density field in large void regions which subsequently perceives noise as a dominant feature. The detections marked in white show high-density regions in which filamentary detections are disjointed since the area is quite dense such that even significant filamentary structures will be labelled as insignificant compared to its surroundings.

Figure 3-8, as well as the information provided in this section help illustrate that although persistence has trouble in delineating tenuous structures from noise due to its topological approach, it provides an extremely successful mathematical tool in simplifying the data set.

3.3.5 DisPerSE

Discrete Persistent Structure Extractor (DisPerSE for short) implements a scale-free, topological approach when analysing the cosmic web for both observational data and

N-body simulations. Based on discrete Morse theory, it partitions space into a set of ascending and descending d-manifolds allowing for the identification of the different morphological environments. Figure 3-10 illustrates the delineation of the constituents with large empty voids defined as ascending 3-manifolds, walls being thin planar structures correspond to ascending 2-manifolds, filaments ascending 1-manifolds and clusters ascending 0-manifolds when using discrete Morse theory. To reduce the noise and provide reliable results, DisPerSE uses persistent homology to look at the k-cycle lifetime of critical pairs. The mathematical mechanisms described in the previous sections allow DisPerSE to extract topological properties of the environment in a rigorous fashion that doesn't necessitate an arbitrarily chosen user-defined scale.

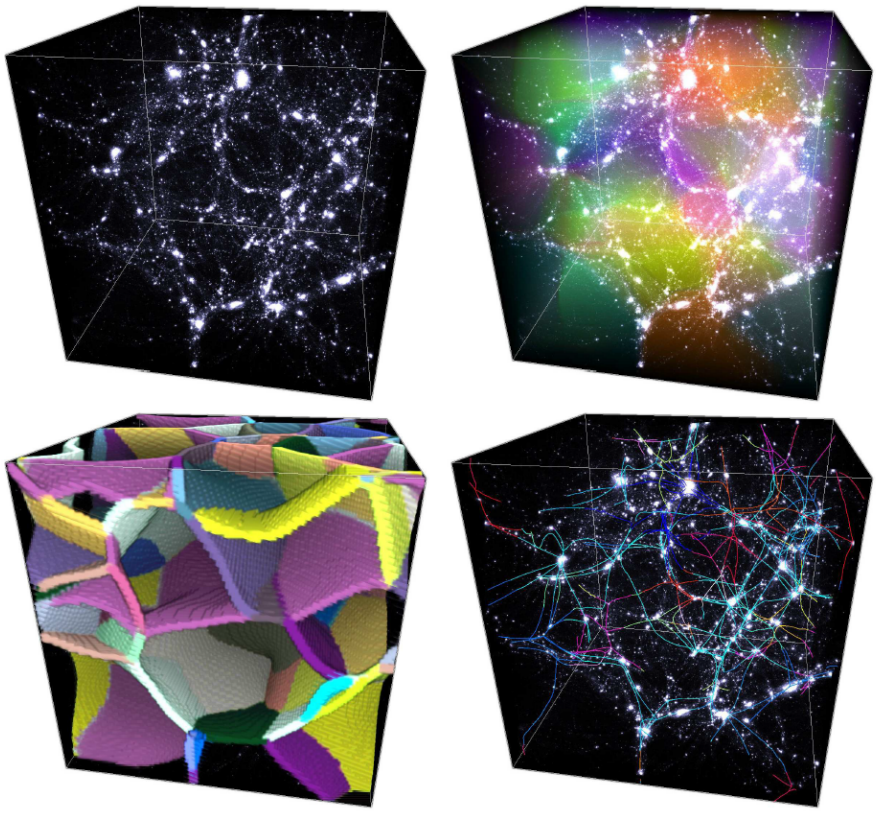


Figure 3-10: For a given dark matter distribution in a simulation (shown by the top-left frame), the top right frame shows the ascending 3-manifolds (voids), the bottom left frame shows walls (ascending 2-manifolds) while the skeleton of the filaments (ascending 1-manifolds) are shown in the right bottom frame. (image courtesy of Sousbie (2011))

It is possible to summarise the DisPerSE formalism through four steps (Sousbie, 2011):

1. After the reconstruction of the density field via DTFE, every point is given a certain density value of ρ. A discrete Morse function is then heuristically defined over the complex with particular weight values given to each d-dimensional simplex.

2. The algorithm proceeds to compute the discrete gradient from the given function to form a discrete Morse-Smale complex.

3. Analysis of the topological and geometrical properties of the given density field is done via the discrete Morse-Smale complex. Critical points overlapping the ascending and descending manifolds are identified between the different environments.

4. Based on the user-given persistence level $nsig$, filtration of the discrete Morse-Smale complex occurs by removing any pair below the given ratio. Both artificial noise and topologically insignificant structure are then robustly removed.

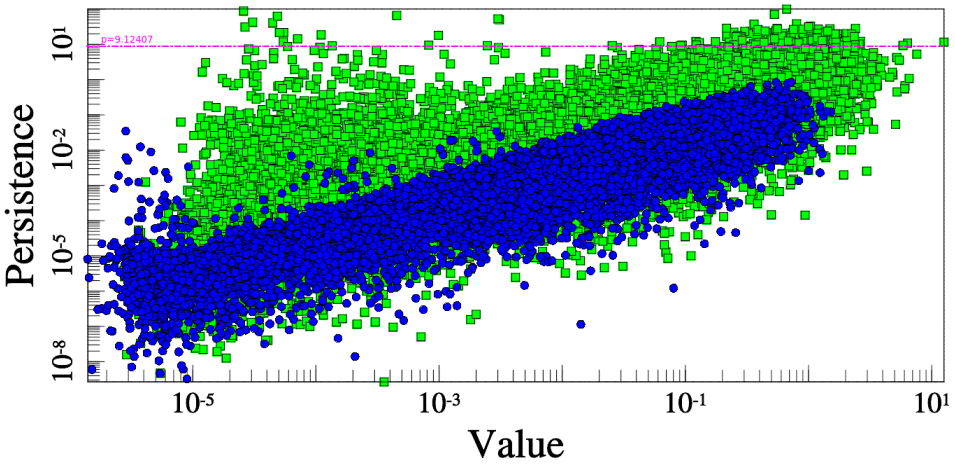


Figure 3-11: DisPerSE provides users with an interactive way when choosing the persistence threshold. The x-axis shows the smaller valued critical point between the pair, the background density. The y-axis denotes the persistence of the pair, or the difference in the discrete Morse function value taken at the specified point.

DisPerSE provides the user with an interactive way of selecting a filtration level in the form of a plot shown in figure 3-11. The plot above shows a persistence pair's lowest value in the x-axis, entailing that as you move along the x-axis, the denser the local environment is. The y-axis denotes the persistence of the pair, such that as you go up the y-axis, the more structurally significant the structure is.

When choosing a persistence level, the user wants to select a value in which the cut off provides only a few distinct points. This value is chosen by control-clicking the plot at the y coordinate you want, as can be seen in pink with a $nsig = 9.12$ value. This value is too large as we see a distinction between features at lower values with green squares representing a saddle/maxima pair start to disperse at earlier ratios.

This figure also illustrates the excess amount of features obtained when using low persistence ratios. The sheer amount of green squares and blue circles denoting minima's (or voids) is staggering and only supports the notion of persistence homology being essential for a mathematically coherent trimming of the network to occur.

In general, DisPerSE identifies the critical points of the density field via Morse theory and does so before classifying whether they are minima's, maxima's or saddle points. By tracing the intrinsic topological features of a given field and attributing them as features within the different environments of the cosmic web, DisPerSE gives a concrete mathematical equivalent to each environment making it commonly used by researchers upon their investigation of filaments (Schneider et al. (2013); Ho et al. (2018); Bonjean et al. (2019); Kraljic et al. (2020)). Having now taken the time to gain the mathematical knowledge on understanding the mechanism behind DisPerSE, it is now time to dissect the NEXUS+ algorithm.

3.4 Scale-Space, Geometric Approach

Mathematically less rigorous than DisPerSE, NEXUS+ investigates the cosmic web using a scale-space analysis to extract the geometrical information ingrained in the underlying density field. This scale-space approach NEXUS+ adopts allows for the tracing of the substructures present within the environment and allows for the identification of the more tenuous structures.

3.4.1 Hessian Density Matrix

The cosmic web theory states that different morphological environments can be distinguished based on the eigenvalues of the tidal field (Bond et al., 1996). Similarly, the density field and more precisely, the Hessian matrix, can be used to delineate the different morphological constituents. The Hessian matrix of the density field forms the basis of NEXUS+, and although DisPerSE uses it as well to distinguish the critical points of the Morse function, it is more apt to place it within the NEXUS+ discussion due to its more pronounced link to it.

The Hessian matrix is a square matrix comprising of second-order partial derivatives, describing a particular field. Mathematically this amounts to:

$$\mathcal{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \tag{1}$$

where the scalar field, f, is a user-inputted density field constructed from a particle sample with the help of DTFE that works for both observational data as well as N-body simulations.

The density field and its local geometry allows NEXUS+ to detect and preserve the shape and size of the different environments since Taylor expanding the density field around a point and conserving second-order terms, information on the local geometry of a point is obtained. Hahn et al. (2007) expanded on this by proposing that the eigenvalues of the Hessian matrix determine the structures of the cosmic web based on specific constraints, shown in table 1, with voids being the unidentified regions.

Structure	Eigenvalue Ratios	Eigenvalue Constraint
Cluster	$\lambda_1 \approx \lambda_2 \approx \lambda_3$	$\lambda_1 < 0, \lambda_2 < 0, \lambda_3 < 0$
Filament	$\lambda_1 \approx \lambda_2 >> \lambda_3$	$\lambda_1 > 0, \lambda_2 < 0, \lambda_3 < 0$
Wall	$\lambda_1 >> \lambda_2 \approx \lambda_3$	$\lambda_1 > 0, \lambda_2 > 0, \lambda_3 < 0$

Table 1: The different morphological environments and their corresponding eigenvalues based on Hahn et al. (2007)

One can understand the link between the eigenvalues of the Hessian matrix and the different environments by imagining a relief map of the density field which has bumps signifying

areas with densities larger than the cosmic mean density, and crevices, under-dense regions with $\rho/\bar{\rho} < 1$ or densities lower than the cosmic mean density of the underlying field. By following the signs of second derivatives, a negative second-order derivative tells you that the function is approaching a maximum. Since clusters are the densest environment within the cosmic web, they become local maxima's on a density relief map. This idea helps illustrate why and how eigenvalues of the Hessian density matrix form the basis of the NEXUS+ formalism since they provide information on the local geometry of the density field allowing for identification of the different environments.

In general, the Hessian provides direct information on the local geometry of the field as it encodes the local curvature of the density field and therefore by using it NEXUS+ renders results that preserve the shape and structure of the identified environment in a user-free way giving for accurately rendered results. By computing the local eigenvalues which are dependent on the local geometry of the field as well as its spatial distribution, the different morphological structures can be identified (Aragón-Calvo et al., 2010). Although a second-order Taylor expansion is used as well as identifying environments purely based on information encoded within the density field, by adopting a scale-space analysis, NEXUS+ bypasses the limitations on its identification technique by examining the substructures present in the cosmic web such that it can probe the connectivity and intricate nature of the cosmic web.

3.4.2 Scale-space Analysis

The scale-space analysis was initially used in medicine to investigate walls and blood vessels (Frangi et al. (1998); Sato et al. (1998)) and was later adapted by Aragón-Calvo et al. (2010) in the Multiscale Morphology Formalism (hereafter MMF) to investigate the cosmic web. By incorporating the eigenvalues of the Hessian density matrix for a range of scales, the MMF examines the substructures present and examines the connectivity between the environments, highlighting the hierarchical and multi-scale nature of the cosmic web.

Upon computing the eigenvalues of the underlying density fields' Hessian matrix for a range of resolutions, the MMF determines the environment for every point based on the criteria's given in table 1 as well as certain physically motivated thresholds. Since eigenvalues are related to the local environment of the field when changing the resolution the local neighbourhood, the environments signal strength also changes allowing for probing of the substructures present within the cosmic web. NEXUS+ implements this formalism using a log-Gaussian density filter and the scale-space analysis MMF provides is summarised with the following (for more details see Aragón-Calvo et al. (2010)):

To start, the DTFE constructed density field, f_{DTFE}, is smoothened over a range of log-Gaussian filters W_G with widths R_n such that:

$$f_n(\vec{x}) = \int f_{\mathrm{DTFE}}(\vec{y}) W_G(\vec{x}, \vec{y}) d\vec{y} \tag{2}$$

where

$$R_n = \sqrt{2^n} R_0 \tag{3}$$

The $\sqrt{2^n}$ ratio is a value taken from Sato et al. (1998) and introduces a 4% deviation with respect to the idealised case of a continuous level analysis. After filtering many different resolutions, a scale-space is constructed by combining the different density maps f_n, or mathematically:

$$\Phi = \bigcup_n f_n \tag{4}$$

Where each point and its corresponding feature is identified via the maximum morphological signal strength, $\mathcal{S}$, it's received. The morphological signal strength is determined based on how well the local neighbourhood of the point correlates to the eigenvalue constraints and ratios indicated in table 1. NEXUS+ implements a soft filtration method of these points based on the eigenvalue ratios since this determines the shape of the local environment. For instance, a cluster region should be somewhat spherical in shape hence why all three of its eigenvalues should be roughly the same, whereas filaments, who have λ_1 and λ_2 opposite in signs but similar values, should be regions collapsed along two if its axis. A more rigorous, physically motivated filtration also takes place and takes into account either the change in the total mass of a given environments population as a function of increasing threshold or via another physically motivated criterion.

The threshold for clusters looks to eliminate small cluster structures while still preserving the large virialised ones. To determine whether an object is virialised, the average density of individual clusters must exceed $\Delta = 370$, a value based on the spherical collapse model at a redshift of zero (Gunn and Gott, 1972). The signature threshold is taken as the value in which half of the detections have a density larger than Δ (Cautun et al., 2013).

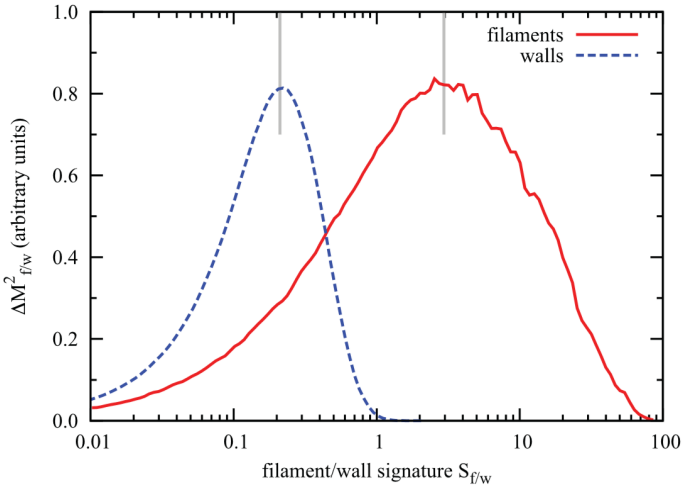


Figure 3-12: Figure showing the change in mass of filaments or walls as a function of significance threshold. (image courtesy of Cautun et al. (2013))

For filaments and walls, the threshold is different. Let $M_f(\mathcal{S}_f)$ denote the total mass of all the detected filaments with a signature value equal to or larger than $\mathcal{S}_f$. As one decreases this signature value, the population constituting the filamentary network grows and therefore the square of the total population mass M_f^2 and the square of the difference in mass, ΔM_f^2, will increase. At a certain point, however, decreasing $\mathcal{S}_f$ causes most of the new filaments entering the subset to originate from sparse or insignificant structures and the value ΔM_f^2 starts to plateau. This turning point is the threshold used by NEXUS+ as it illustrates the limit between the detection of significant and spurious structures. Figure 3-12 helps illustrate this idea.

The biggest advantage scale-space analysis has in analysing the cosmic web is via the identification of the different structures in a user-free manner that allows for probing of the hierarchical nature of the cosmic web, giving it the ability to adapt and investigate the cosmic web as well as its multi-scale nature in a coherent fashion. It allows probing of the multi-scale nature since by using a range of scales, NEXUS+ can extract information on the more tenuous of environments as well as the significant ones. Furthermore, since walls are sparse by nature, it is hard to detect them and therefore are not always identified in a single scale analysis. By using a range of scales, NEXUS+ is also able to render wall detections successfully.

3.4.3 A Smoothing Scale Comparison

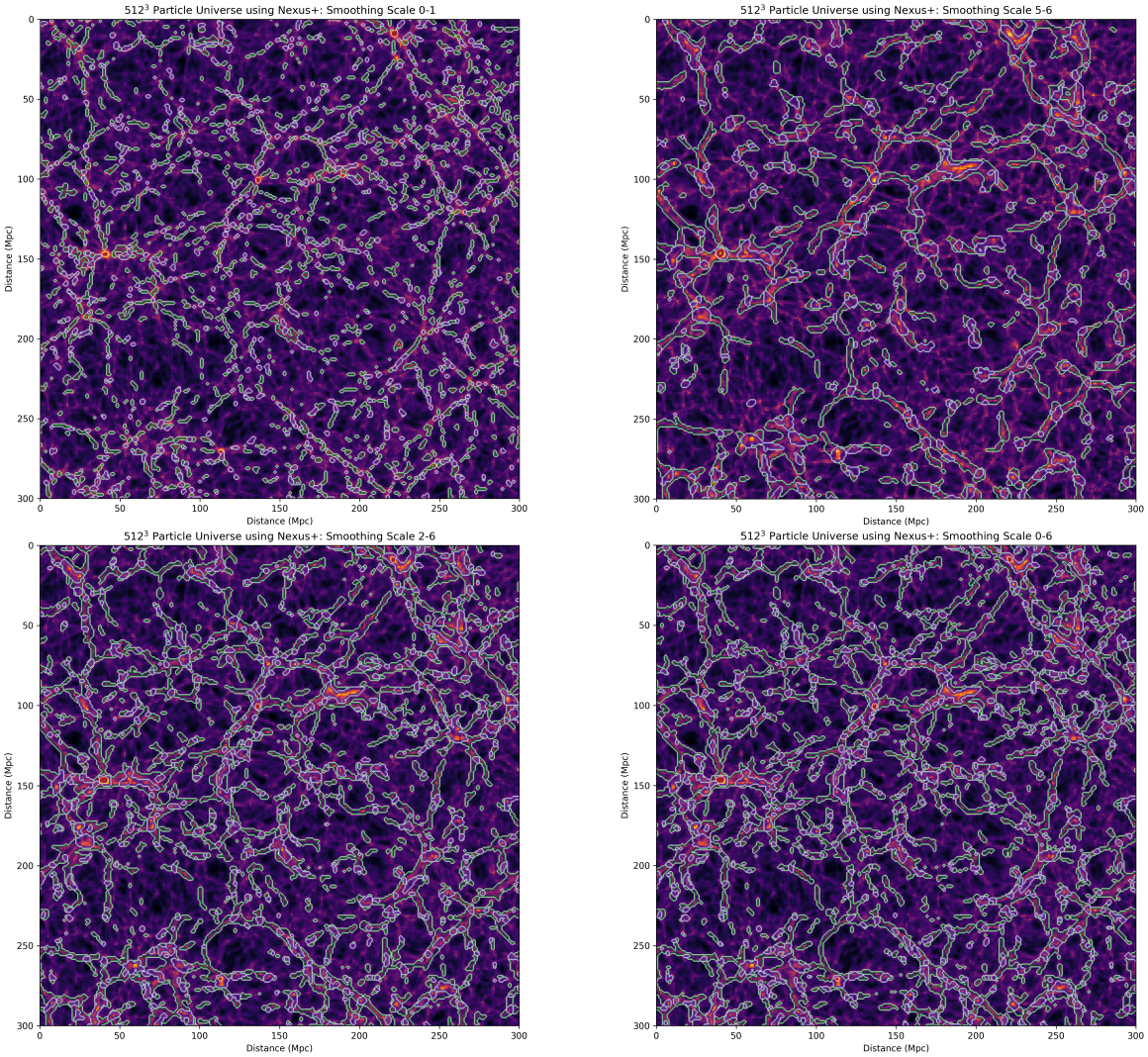


Figure 3-13: Identification of different morphological environments conducted on a 512^3 particle ΛCDM universe for a range of scales. The density field is visualised in colours where red, bright regions signify dense regions while vast, empty dark regions voids. The green contour lines signify walls while purple contour lines filaments. A few clusters are identified and highlighted with red contour lines. Top left: The Universe as analysed by a scale ranging from $0 \leq n \leq 1$. Top right: Scale analysis of the Universe at scales ranging from $5 \leq n \leq 6$. Bottom left: Scale analysis of the Universe at scales $0 \leq n \leq 0$. Bottom Right: Scale analysis of the simulated Universe using the scales $2 \leq n \leq 6$.

Figure 3-13 shows four samples of the same Universe each analysed at different scales in which $R_0 = 2h^{-1}$ Mpc and $0 \leq n \leq 6$. This range of n corresponds to $R_n = [2.00, 2.83, 4.00, 5.66, 8.00, 11.3, 16.0]h^{-1}$ Mpc smoothing filters following equation 6. The top two frames show the Universe analysed over two scales and the lower ones consist of an analysis using a larger range of scales. One observes that the larger filter values ($5 \leq n \leq 6$) tend to identify larger and thicker environments, correlating to significant

structures. At the other end of the spectrum, smaller filters pick out the tenuous structures with thin, delicate purple filament and green wall tendrils being identified. These smaller scales have the limitation that noise becomes more dominant a feature, giving rise to errors if not taken into consideration and highlights the need for a scale-space approach when analysing the cosmic web.

The lower two frames show the identification of morphological components corresponding to smoothing scales with $0 \leq n \leq 6$ and $2 \leq n \leq 6$. Upon comparison of the two, one sees that by using a larger range of scales, NEXUS+ identifies more structures. These detections, however, don't necessarily correlate to significant environments and instead tend to be unaccounted for structures in void regions further showing that at the smaller scales resolution effects come into play as noise becomes more dominant a feature or haloes being miss identified as walls or filaments due to their densities. We see that the detections corresponding to $2 \leq n \leq 6$ remove these features by not taking into account these smaller scales.

One also notices from the bottom two frames that by using a range of different scales NEXUS+ can delineate fragile tendrils of walls and filamentary regions while still incorporating the large, significant structures found by the larger scales. A combination of a range of smoothing scales provides not only extraction of information on both these fragile environments as well as significant ones but also allows for a better rendering of the shapes and sizes of the different environments in a natural, unbiased fashion.

A more quantifiable comparison between the different smoothing scales are seen in the two bar graphs below (Appendix E provides the individual pie charts).

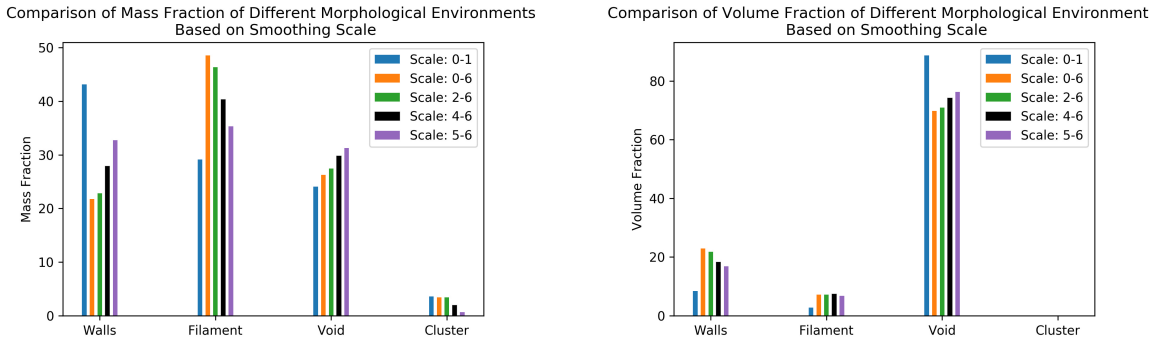


Figure 3-14: Bar graphs illustrating the different mass fractions (left frame) and volume fractions (right frame) each environment occupies within the Universe based on the identification carried out at different smoothing scales.

The figures above show the different mass and volume fractions of the simulated Universe. By making up the dominant fraction of the mass component of the Universe, filaments are the most prominent feature within the cosmic web. Clusters, on the other hand, even though they are the densest environment within the cosmic web, occupy a negligible fraction of the volume making its mass composition insignificant at first order approximation. On

the other hand, voids dominate the volume content of the Universe and account for a significant portion of the mass fraction of the Universe even if, by definition, they make up the sparsest regions of the cosmic web.

When examining the bars corresponding to the smoothing scale of $0 \leq n \leq 1$ and $5 \leq n \leq 6$ compared to those of $0 \leq n \leq 6$ and $2 \leq n \leq 6$, we see that the smaller the range of scales used, the larger the mass bias towards walls is at the detriment of the filaments. This trend stems from the fact that filaments are elongated structures, collapsed along two axis and occupy less volume than the sheet-like walls. Without a range of scales to identify the structure, the filter may skip over an arbitrary resolution which could otherwise provide a strong significance $\mathcal{S}$ towards filaments. It's easier to skip over an essential scale for filaments compared to walls since they have undergone collapse upon two of their axis and also occupy a smaller volume fraction of the Universe. This observation provides insight as to why a range of scale for the analysis of the cosmic web is necessary since it allows for better detection of the different environments without skipping essential scales as well as being able to analyse both tenuous and significant structures.

Furthermore, the $0 \leq n \leq 1$ scale has a large discrepancy between the volume and mass fractions occupied by both walls and voids. This difference for walls stems from the detection of noise and tenuous filaments being incorrectly given a morphological signal corresponding to walls. By adding the noise and filament detections, the mass fraction of walls will increase without the volume increasing as much. Conversely, the difference between the volume and mass fractions of voids compared to the other scales stems from smaller scales analysing smaller slices of the Universe. By analysing smaller slices without looking at the bigger picture, NEXUS+ ignores the multi-scale nature of the cosmic web, subsequently making it identify most of the tenuous regions as voids. Figure 2-1 supports this by showing that the overall cosmic web and void PDF are similar, leading to small scales detecting most regions as voids since they are unable to look at the bigger picture and base their signatures by these larger scales.

The smallest scale, however, does provide for analysis on clusters as well as the intricacies of the cosmic web. Clusters are dense, compact regions of space and since a smoothing scale can only analyse environments roughly its size or larger, clusters will only be detected using small scales, since at larger scales their densities will appear to merge with the surrounding regions. This also supports the notion that smaller scales will identify noise as structurally significant structures whereas larger filters will skip it along with the tenuous tendrils of the filamentary and wall network, leading us to conclude that a wide range of scales is essential for the investigation of the cosmic web.

By analysing the Hessian matrix of the density field for a range of resolutions, MMF explicitly allows for a scale-space analysis of the cosmic web making for an essential formalism that NEXUS+ can expand on. Aragón-Calvo et al. (2007) showed that by using the MMF, Poisson noise induced by the density field had less of an effect on the detection of different environments. This all amounts to a successful tracing of the asymmetrical

pattern exhibited by filaments and walls of the cosmic web as well as its successful probing of the hierarchical nature of the cosmic web with the major downside being the lack of rigorous mathematical approach when identifying the environments.

3.4.4 NEXUS+

NEXUS is a formalism which encompasses several different tracers ranging from the tidal field to the density field to various velocity fields, be it divergence or shear. Though the tidal field provides information on the nature of the anisotropic gravitational collapse and therefore essential information about the mechanisms behind structure formation, the sharp contrast in densities between various environments within the cosmic web makes the density field an ideal choice for its analysis.

The NEXUS formalism has two different density-based tracers. One uses a simple density field while the other, NEXUS+, looks at the environment using a log-Gaussian filter and will be the focal points for this report. Cautun et al. (2013) showed that by using a log-Gaussian filter, large scale structure of the Universe were rendered better. Furthermore, since the NEXUS formalism works more efficiently upon comparing structures of similar values (Cautun et al., 2013) and since different morphological environments can span many orders of magnitudes in density, by taking the logarithm of the density field the contrast in densities within the cosmic web only differ by a few integers giving for better results. This logarithmic density field allows NEXUS+ to be sensitive to both prominent, as well as tenuous structures of the cosmic web, with the downside being the loss of information on the large-scale structures around density peaks since their contributions are reduced.

NEXUS+ identifies the different constituents of the cosmic web by following six steps (for a more thorough overview see Cautun et al. (2013)):

1. After constructing the density field $f(x)$ via DTFE, the logarithm of the field is taken such that $g = \log_{10}(f)$. In doing so, a Gaussian filter W_G is applied giving for a smoothed density field $g_{R_n}(x)$ through solving:

$$g_{R_n}(x) = \int g(y) W_{G,R_n}(x, y) d^3y \tag{5}$$

   In general this is done via the convolution of $g(y)$ with that of the Gaussian filter in Fourier space such that the integral becomes:

$$g_{R_n}(x) = \int \frac{1}{(2\pi)^3} \hat{g}(k) e^{ikx} e^{-\frac{k^2 R_n^2}{2}} d^3k \tag{6}$$

   where $\hat{g}(k)$ is the Fourier field component and $e^{-\frac{k^2 R_n^2}{2}}$ the Gaussian exponential.

2. By processing this filter throughout the logarithm of the density field the Hessian matrix $\mathcal{H}$ becomes:

$$\mathcal{H}_{ij,R_n}(x) = R_n^2 \frac{\partial^2 f_{R_n}(x)}{\partial x_i \partial x_j} \tag{7}$$

and the Fourier transform;

$$\mathcal{H}_{ij,R_n}(x) = -k_i k_j R_n^2 \hat{f}_{R_n}(k) \tag{8}$$

Multiplying by R_n^2 allows the terms in the Hessian to get renormalised such that there is no bias between scales (Cautun et al., 2013).

3. After filtration of the Hessian matrix, the signs of the eigenvalues and their values are analysed to correlate regions with environments based on the criteria's shown in table 1, the stronger it correlates to a particular environment, the larger it's signature $\mathcal{S}$ will be. By placing this criterion, NEXUS+ gives an unbiased characterisation of the different environments with respect to their shape or size.

4. NEXUS+ then repeats the previous three steps for a range of scales chosen by the user in increments of $R_n = \sqrt{2^n} R_0$, giving for a different morphological signature $\mathcal{S}$ for each point at each smoothened scale.

5. By combining the results for all the scales, NEXUS+ obtains a multi-scale, parameter-free analysis on the different environments based on their eigenvalues. For every point it will assign the morphological environment based on the maximum morphological signature the point was assigned through the range of scales:

$$\mathcal{S}(x) = \max_{\text{levels } n} \mathcal{S}_{R_n}(x) \tag{9}$$

6. The final step applies the clusters as well as walls/filaments thresholds that were briefly discussed in section 3.4.2. By filtering through the signatures based on these physical criterion, noise gets further eliminated giving for more reliable results.

In his study, Cautun et al. (2014) investigated the cosmic web using all the different NEXUS formalisms and found that the mass and volume fraction between one another differed by $< 10\%$, this suggests that its thresholding is physically motivated and does not depend on the underlying field.

Many studies have used NEXUS+ in its analysis of the cosmic web (Hellwing (2016); Ganeshaiah Veena et al. (2018); Ganeshaiah Veena et al. (2019)). In the paper that introduced the formalism (Cautun et al., 2013), it's found that by adopting a scale-space approach in analysing the cosmic web, NEXUS+ allowed for better detection of the tenuous structures while still preserving the filamentary and wall networks of the cosmic web. Hellwing (2016) reiterated this in his paper by showing that due to the discrepancy between the dispersion of different morphological environments, NEXUS+ provides a robust algorithm in delineating the different constituents.

From the steps mentioned above, it is clear that NEXUS+ provides a scale-space analysis of the cosmic web and on paper seems to agree with the attributes of what makes a cosmic web tracing algorithm successful, mentioned in section 3.1. It gives a direct comparison between environments by looking at the eigenvalues for a range of scales, allowing for coherent

information on the connectivity between different environments to be extracted. Section 3.4.2 illustrated its ability in finding the more tenuous structures as well as successfully delineating the more significant ones, and does so by using a log-Gaussian filter on the input density field. Bonnaire et al. (2019) provides a summary on the formalism by describing it as an algorithm that provides a scale-space representation of the density field, where different environments are defined by their corresponding eigenvalues of the Hessian density matrix obtained through the DTFE method.

3.4.5 NEXUS+ vs. NEXUS_den

As stated in the previous section, NEXUS incorporates different density tracers. Though both methods generally follow the same algorithm, NEXUS+ examines the underlying density field with a range of log-Gaussian smoothing filters, while NEXUS_den incorporates a linear Gaussian filter, causing some discrepancies in their results.

As we have seen, the cosmic web encompasses many orders of magnitudes of densities. Since NEXUS is more efficient when structures have similar values, a log-Gaussian filter will work better since the density field will have a narrower range in densities and this, in turn, will allow detection of tenuous features to occur.

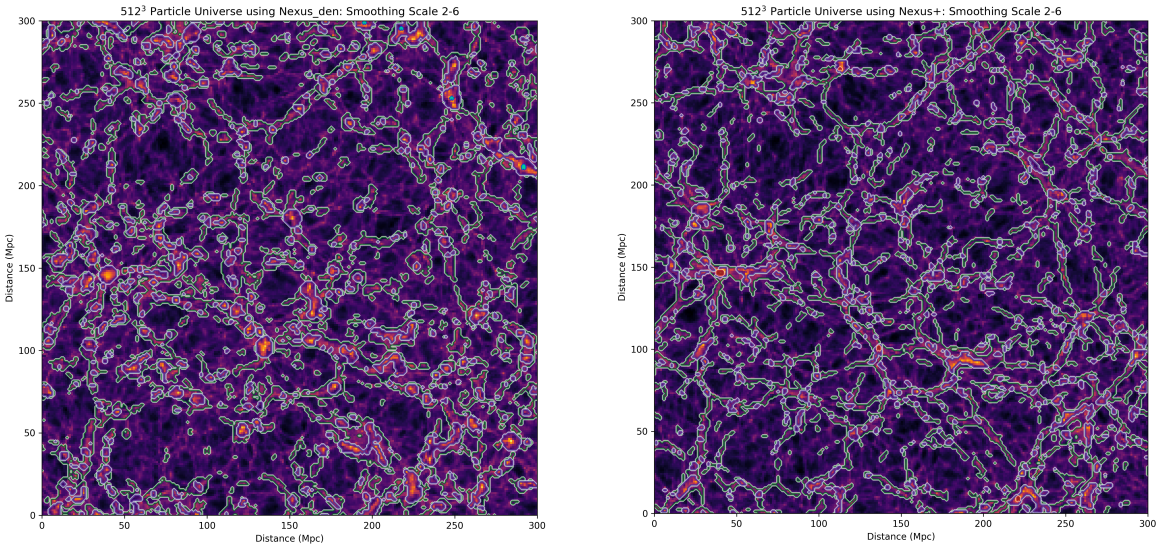


Figure 3-15: Left Frame: The cosmic webs filament and wall detections using NEXUS_den. Right Frame: The cosmic webs filament and wall detections when using NEXUS+

Although NEXUS+ has more success when tracing the anisotropic structures of the cosmic web, it struggles when detecting clusters. Since clusters are dense regions of space, by using a log-Gaussian filter, the contribution of a highly dense core will be minimised, therefore having more chance of being left undetected. For this reason, the report will utilise NEXUS_den when examining the cluster regions.

Straight away one observes from figure 3-15 that both algorithms trace the significant features of both walls and filaments well. Nonetheless, the filamentary network found by NEXUS+ is much thinner and smoother than the one detected by NEXUS_den. Since NEXUS_den incorporates a linear Gaussian filter, a large range in values will cause detected environments to smudge itself over a wider region since densities will need to differ by a larger amount if it wants to be identified as a different environment. This causes identified regions to absorb both underdense regions (as seen in figure 3-15 with the long wall and filament detection at $(x, y) = (260, 250)$ as well as cluster regions. When examining figure 3-15, one notices that in dense regions, NEXUS_den identifies circular-shaped filaments contrary to the elongated ones found by NEXUS+. The detection found by NEXUS+ better correlates to the shape one would expect from an environment collapsed along with two of its axis while cluster regions illustrate perfectly the smudging effect that occurs when using NEXUS_den. Since cluster regions are dense, compact regions of space assumed to be spherical in shape, the outskirts of a cluster region will induce significant smudging of a detected environment, manifesting itself in a circular form as observed in the figures.

Figure 3-16 highlights the success NEXUS+ has in identifying the tenuous filaments as most of its uncommon detections lie on the outskirts of the network in the shape of thin tendrils whom branch off from the main filamentary network and tending towards sparser regions. The fact that NEXUS_den is unable to pick out the tenuous features of the cosmic web suggests that NEXUS+ allows for a better analysis of the connectivity between different environments.

The two frames in figure 3-17 show zoomed-in slices at certain sections of the field. On the left-hand side, we see a filamentary network, on the right-hand side the identified walls. The figure further illustrates the success NEXUS+ has in the identification of the more fragile structures of the cosmic web since it shows features lying in void-like regions branching off from the network. Though it is hard to conclude with certainty whether they are genuine features, with the knowledge that the NEXUS formalism works better when values incorporate a narrower range of densities as well as the thin detections NEXUS+ finds with regards to filaments and walls branching off the network, one concludes that NEXUS+ renders better results which subsequently manifest themselves in the quantitative

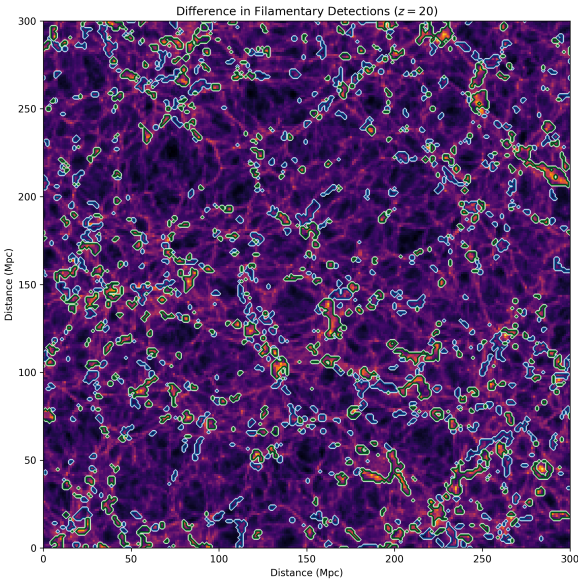


Figure 3-16: The differences in the detection of filaments. Blue contour lines denote uncommon detections found by NEXUS+ while green contour lines denote the uncommon detections found using NEXUS_den.

results extracted.

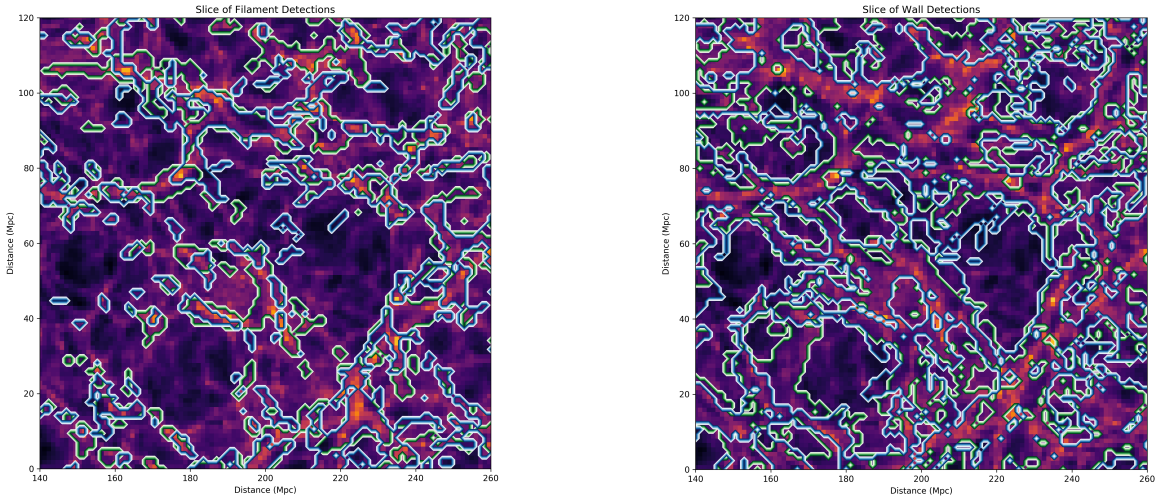


Figure 3-17: Left Frame: A zoomed-in slice of detected filaments where green denotes identified structures using NEXUS_den and blue contour lines those of NEXUS+. Right Frame: A zoomed-in slice of the wall network as detected by NEXUS_den in green and NEXUS+ in blue.

The differences in the identified wall network are similar to the ones found between their filamentary network. In most cases, the walls found using NEXUS_den envelope those of NEXUS+. Again, one observes that NEXUS+ identifies tenuous or fragile wall features branching off into void regions in the form of delicate tendrils (i.e at the void $(x, y) = (230, 50)$). The fact that these detections are thin and branch off from dense regions to underdense ones suggests that NEXUS+ is better at delineating tenuous features. If these detections were thick such as the one found at $(x, y) = (170, 30)$ one could assume a large fraction of error. Furthermore, gaps appear in the structure identified by NEXUS_den. These gaps are due to walls now and again consisting of dense halos. Due to its linear analysis of the cosmic web, NEXUS_den will identify these dense halo regions as non-wall like but rather as filaments or clusters.

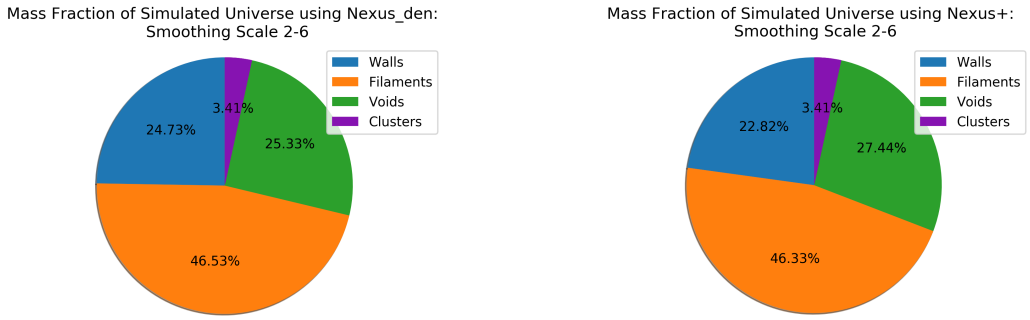


Figure 3-18: The mass fraction of the different environments as detected by NEXUS_den (left) and NEXUS+ (right)

The mass fraction shown in figure 3-18 is similar between both algorithms. Nevertheless, a few discrepancies emerge due to the qualitative effects discussed earlier. The smudging of dense regions into underdense ones manifests themselves in the mass portion of voids with the one detected by NEXUS_den being roughly 2% larger than that found by NEXUS+. With voids identified as the remaining environment, its mass fraction illustrates the discrepancy in the overall detection of both algorithms. By detecting wider environments, void regions will inevitably be absorbed, causing the mass attributed to voids via NEXUS_den to decrease at the benefit of both walls and filaments.

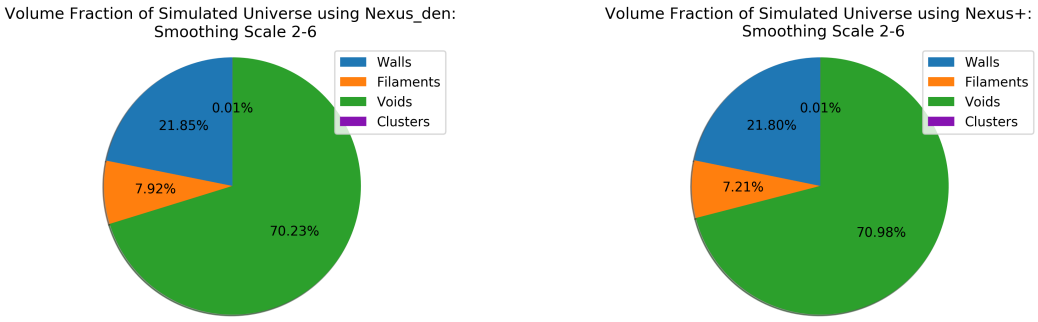


Figure 3-19: The volume fraction of the different environments as detected by NEXUS_den (left) and NEXUS+ (right)

The reason why filaments gain less mass than those of walls is that filaments are denser than walls and so when smearing occurs, underdense regions will be falsely identified as filaments leaving for a smaller effect on its mass, whereas for walls there's more chance that denser regions will be absorbed giving it a more pronounced mass increase. This effect manifests itself in the volume and mass fraction of both Universe as seen in figure 3-18 and 3-19 where the mass fraction of walls exhibit a large difference, whereas, for filaments, the difference in volume becomes more pronounced between the two algorithms.

Cumulative distribution function (CDF) plots are good at underscoring the distribution of detections for different algorithm and helps illustrate the discrepancy in tenuous detections between the two algorithms.

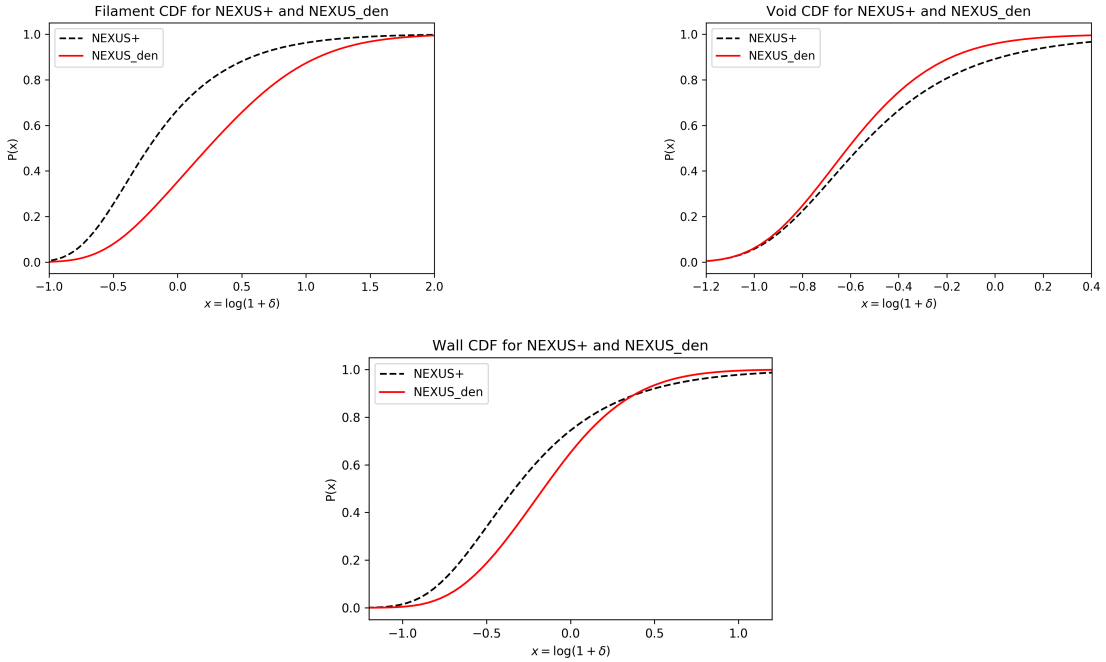


Figure 3-20: The cumulative distribution function of the different environments density as found by NEXUS_den and NEXUS+.

The three plots all show that the log-Gaussian approach taken by NEXUS+ has more success in detecting the tenuous features of the cosmic web for both walls and filaments since it incorporates a higher percentage of the total identified population. As alluded to, this causes the remaining environment, voids, to have less of structurally significant identifications explaining the reason as to why the distribution seems to flip for the CDF of voids.

Particularly noteworthy is that there are consistently more significant filamentary detections made by NEXUS_den as opposed to tenuous ones, but the same cannot be said about walls with the two trend lines intersecting. NEXUS_den has a large portion of its detections at a section near the upper end of the densities due to walls enveloping filaments and so upon using a linear Gaussian smoothing, the smudging will cause walls to overlap with filaments and therefore the tenuous boundaries of filamentary structures, which tend to be denser than walls, being falsely identified as walls.

The above qualitative and quantitative analysis shows that not only does NEXUS+ give for better rendering of the anisotropic structures, but it also allows for better detection of tenuous features, something essential for studying such a fragile environment. Based on previous papers (Cautun et al. (2013); Cautun et al. (2014)) as well as the results found in this section, this report uses NEXUS+ in the comparison of the anisotropic features and will use NEXUS_den for cluster identification, which only highlights the adaptability the NEXUS formalism provides its user.

4 The Simulation

The particle positions provided by Roi Kugel are based on the Universe simulated by Bos (2016) for his PhD thesis.

4.1 ΛCDM Universe

The Universe is a periodic 256^3 ΛCDM Universe with the following parameters; $\Omega_m = 0.268$, $\Omega_\Lambda = 0.732$, $\Omega_b = 0.044$, $h = 0.704$, $\sigma_8 = 0.776$ and $n = 0.948$ (Bos, 2016). Ω_m and Ω_b signify the mass composition of both dark matter and baryonic matter respectively while Ω_Λ denotes the dark energy composition of the Universe. This model correlates to the current standard model in cosmology and takes into account cold dark matter as well as the cosmological constant, which coincides with observations and results recorded at the largest of scales, for instance, the accelerating expansion rate of the Universe (Riess et al., 1998). For the remaining terms; σ_8 gives the density field fluctuation root mean square value at the $8h^{-1}$ Mpc scale, n represents the power-law index of the primordial power spectrum and h the Hubble parameter of modern-day measurements in scales of 100Mpc km^{-1} s.

Using NEXUS+ on a 256^3 simulation, a minute amount of clusters are identified due to the resolution constraints. However, upon computing a 512^3 sampled Universe, DisPerSE was so computationally intensive using over 150GB worth of memory on the norma supercomputers at the Kapteyn Institute, that it felt more suitable to analyse a 256^3 Universe to give researchers and PhD students priority while keeping in mind that a comparison of both algorithms will not differ too much as long as both simulate the same Universe. Furthermore, the main goal of the report is to investigate the differences in detection of the morphological environments with an ambiguous definition, that is voids, walls and filaments. Due to clusters being virialised environments, we can define them based on a physically motivated definition, and so their analysis isn't essential for the report. Nonetheless, this causes an issue in our comparison as one formalism detects an environment that the other doesn't identify. To fix this, the volume and mass constituents correlating to clusters will be removed from the sample for NEXUS+, while DisPerSE automatically incorporates them into whichever environment it identifies the regions as. Although this difference may change the mass composition of the Universe, the volume composition will remain constant.

In the end, a cube with sides $300h^{-1}$ Mpc representing the density field traced by the dark matter particles are obtained. For an overview, all the parameters of the sampled Universe are given in table 2.

Parameters	Ω_m	Ω_b	Ω_Λ	σ_8	n	h	z
Value	0.268	0.044	0.732	0.776	0.948	0.704	0

Table 2: The parameters of the simulated Universe. All values were taken from Bos (2016).

Finally, as seen by figure 4-1 different slices will have various identifiable features within it. Although the quantifiable information such as density profiles or mass content will use information obtained of the fully sampled Universe, the qualitative analysis uses the projection of a slice at $z = 47$Mpc, a value arbitrarily chosen.

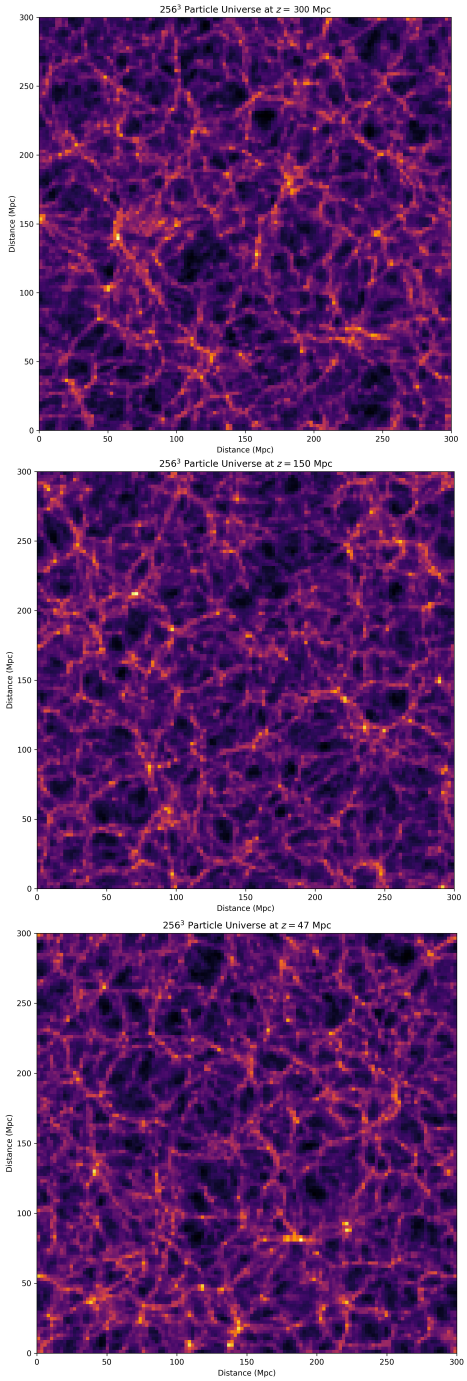


Figure 4-1: Different slices of the sample 256^3 sampled Universe described in section 4.1 all using the same smoothing scale. Top Frame: Slice of the Universe at $z = 300$Mpc. Middle Frame: Slice of the Universe at $z = 150$Mpc. Bottom Frame: Slice of the Universe that will be used during the qualitative analysis, along $z = 47$Mpc

4.2 Persistence Level Parameter

Along with what features the user wants to identify, be it voids, walls or filaments, DisPerSE provides a user-input on the significance level of structures to topologically simplify the underlying field.

As mentioned earlier, a *nsig* level of two gives rise to 5% artificial features while a *nsig* level of four corresponds to an error of 0.006%. Though it is not mathematically supported, the analysis will revolve around a level of $nsig = 3$. This value was chosen based on countless studies using it, much of them had the developer Dr Sousbie himself as help. Furthermore, when trying to use the interactive persistence diagram to choose a better value, the coma machines in the Kapteyn Institute would crash without fail.

4.3 Smoothing Scales

As seen by figure 3-13, different smoothing scales will amount to different environments being identified. For the qualitative and quantitative analysis, a smoothing scale of $R_n \in [4.00, 5.66, 8.00, 11.3, 16.0]h^{-1}$ Mpc or $2 \leq n \leq 6$ will be used. The minimum smoothing scale was selected since going below 2Mpc, or the Nyquist range, resolution effects are introduced in the identification of the environments as noise becomes more dominant a feature for these smaller scales.

4.4 DisPerSE: Tracing of the Filaments

In general, DisPerSE is used to extract the skeleton of the cosmic web, allowing cosmologists to visualise the underlying structure of the network (Appendix F discusses the extraction process further). This extraction of the filamentary skeleton also means that when sampling results on a grid, the environment will be rendered poorly since DisPerSE provides no information on its shape or size.

After processing the field, DisPerSE will provide users with a skeleton file giving coordinates of all points identified towards a particular environment. By determining the connectivity between coordinates, using a 3D software such as ParaView one can observe the underlying structure of the cosmic web as illustrated by the filamentary network in figure 4-2.

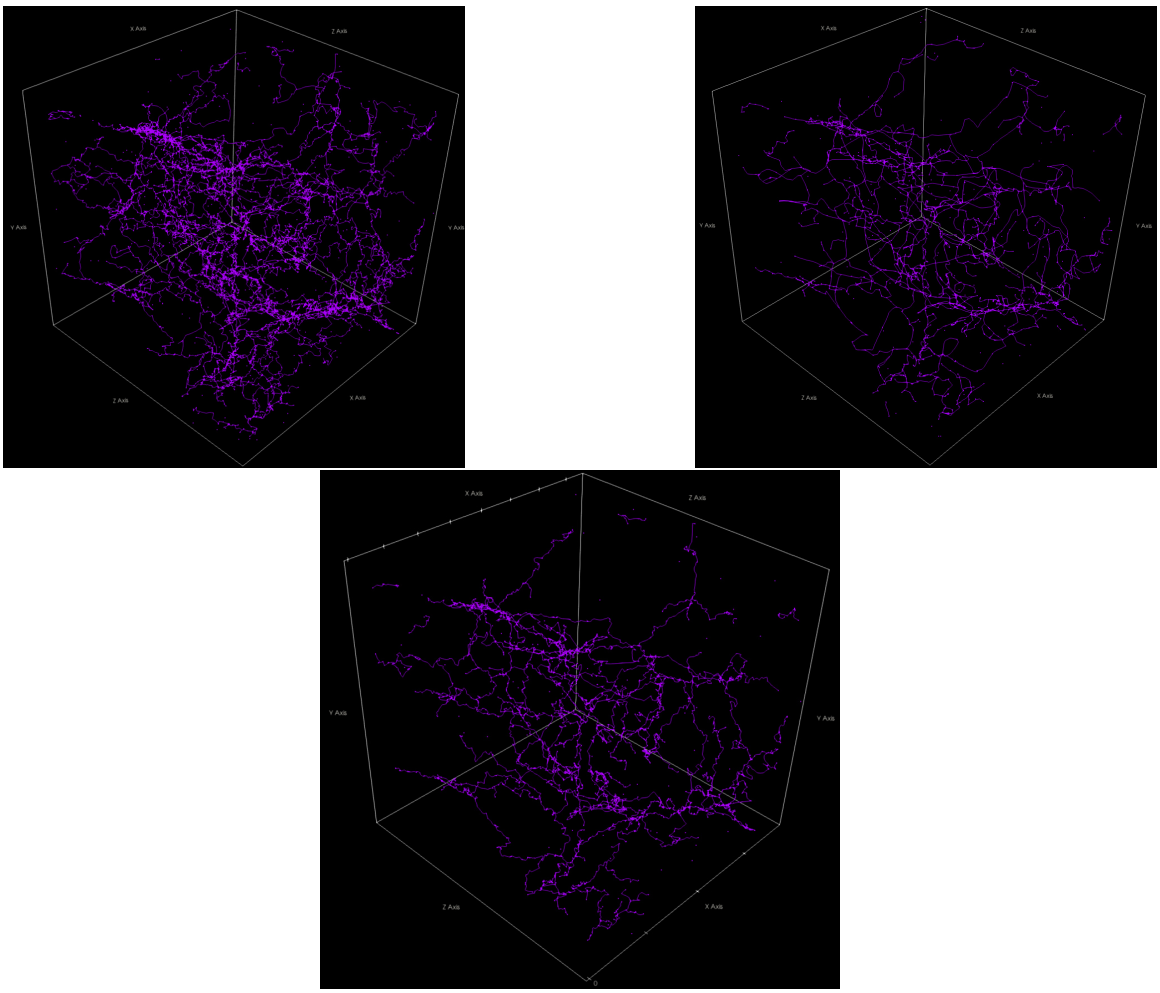


Figure 4-2: The skeleton extraction DisPerSE provides on a $4 \times 4 \times 4$ Mpc^3 slice. Top Left Frame: Persistence level $n\sigma = 2$. Top Right Frame: Persistence level $n\sigma = 3$. Bottom Frame: Persistence level $n\sigma = 4$.

This 3D visualisation of the Universe for different persistent levels not only shows the success persistence has in trimming out noise while still preserving significant structures but also illustrates the strength DisPerSE has in highlighting the connectivity between filaments, making it a successful filamentary extractor.

NEXUS+ identifies environments in a scale-space fashion by evaluating the morphological signal at each cell. NEXUS+ then identifies the morphological environments based on the most significant signal obtained when combining all the resolution maps. This method implies that NEXUS+ is unable to find which structures belong to which network, rather it only identifies environments based on the individual cell.

This method was also selected after consolidating with Dr Kooistra, who has experience with DisPerSE. Dr Kooistra mentioned that he often makes identified cells use a somewhat arbitrary thickness. This thickness can be defined as $2h^{-1}$ Mpc for instance, which coincides to the radius of prominent filaments (Colberg et al. (2005); González and Padilla (2010); Bond et al. (2010)), nonetheless, by implementing a user-defined size for environments systematic errors will arise giving for poor results. For this report, however, this will not be the case as I aim to compare the algorithm in identical fashion such that any differences in results arise purely based on the method they adopt when analysing the cosmic web. Instead, I will sample the extracted coordinates onto a grid with the same resolution as the provided density map, which in turn depends on the number of particles present within the simulation.

5 Results

The results are split into two. The first part compares the qualitative results both formalisms obtain while the second section develops a quantitative analysis focusing on different density, mass and volume properties of the Universe as well as their different filtration methods.

The qualitative results examine the detected features in a 9 Mpc slice unless stated otherwise.

5.1 A Qualitative Comparison

5.1.1 Clusters

Both DisPerSE and NEXUS+ have issues in identifying clusters. DisPerSE cannot delineate the nodes no matter the resolution (Libeskind et al., 2018) and upon using a poorly sampled 256^3 particle Universe, NEXUS+ identifies only a few clusters (highlighted in red under Appendix G) such that no detections were found in the chosen slice the qualitative portion of the results focuses on. Since clusters still make up an essential constituent of the cosmic web, for this section only, a simulation involving 512^3 particles was conducted using the NEXUS_den formalism to gain some insight regarding nodes.

The volume and mass fractions found in section 3.4.3 for the smoothing scale $2 \leq n \leq 6$ are similar to those found by other formalisms (Libeskind et al., 2018), suggesting that at higher resolutions NEXUS_den identifies clusters well. Nonetheless, figure 5-1 shows a few nodes left unidentified by NEXUS_den when we compare it with our perception of the slice.

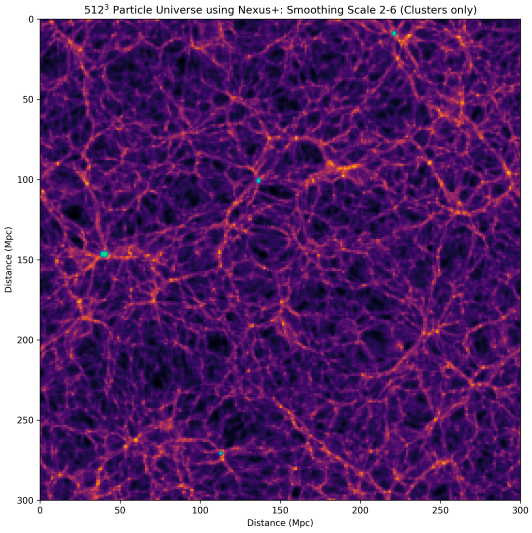


Figure 5-1: Detected clusters by NEXUS+ highlighted in green for a 512^3 particle Universe.

By using the mass and density properties to distinguish if a cluster is virialised or not, The NEXUS formalism adopts a robust method in filtering out cluster populations. The density threshold discussed in section 3.4.2 causes at least half of the clusters to be left unidentified by construct since they are too small and underdeveloped. Moreover, since cluster regions correspond to regions with extremely dense cores, a poorly resolved sample causes the density peak at a point to merge with its surrounding areas, this decreases its core density and causes a change in its morphological significance resulting in the Hessian matrix's eigenvalues to change.

5.1.2 Filaments

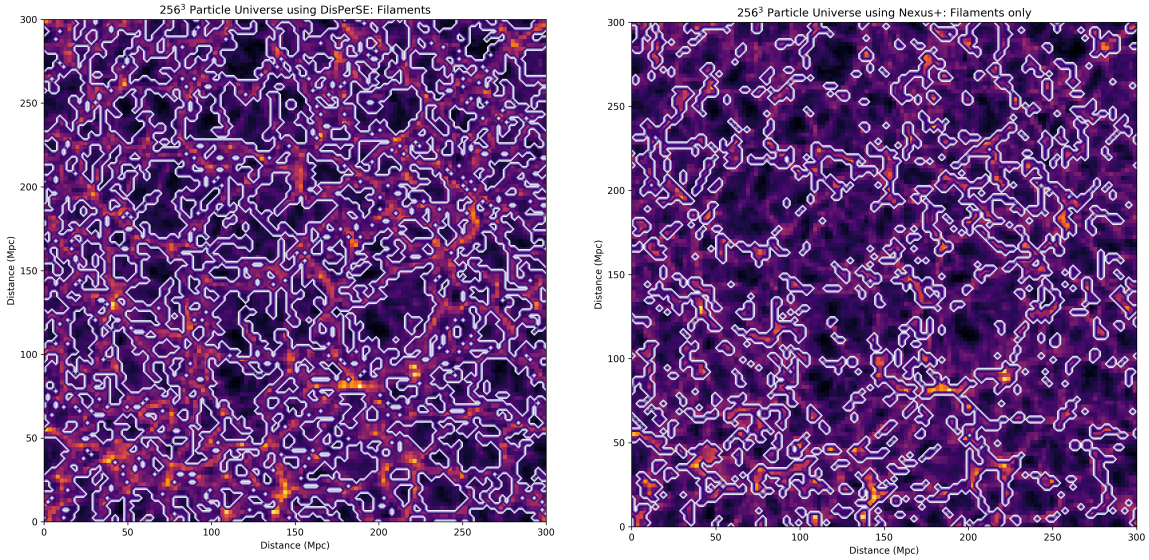


Figure 5-2: Two figures side by side showing the detected filamentary network by both DisPerSE and NEXUS+. The slice is focused at $z = 47$ Mpc but includes detections at a range of $z = [42, 51]$ Mpc Left: Filamentary network as detected by DisPerSE. Right: Filamentary network detected by NEXUS+.

Figure 5-2 shows the filamentary network extracted by both formalisms. Projection effects (better illustrated in Appendix H) cause filaments, an environment that has collapsed along two of its axis, to appear as dots when projected on a two-dimensional plane rather than a surface. Naïvely, one also sees that DisPerSE traces the filamentary network with more success compared to NEXUS+, which has gaps and discontinuity between detections for instance at $(x, y) = (170, 80)$.

One reason why DisPerSE traces the filamentary network nicely is due to its initial assumption that filaments link regions of density maximums via the tracing of its ascending 2-manifold in a discrete Morse-Smale complex. This delineation causes regions between density peaks to be identified as filaments no matter their density as long as it falls within the tracing of the ascending manifold. Although this provides a reliable detection on the significant structures by providing them with a mathematical foundation to fall back on and also allows tracing of the multi-scale nature of the environments, false detections will occur as well as a loss of information on the more tenuous filaments situated in a void or sparser region. At $(x, y) = (250, 130)$ and $(250, 220)$ we see examples of different void regions being detected as filaments. Appendix G visualises the discrepancy in filamentary detections between DisPerSE and NEXUS+, which is also due to the critical points being locally defined, amounting to a surplus of detections on this ill-defined scale.

NEXUS+ does not have this assumption as it provides analysis for all points for a range

of scales based on certain physically motivated constraints defined by the Hessian eigenvalues. This scale-space analysis allows NEXUS+ to investigate the connectivity between the data points for a range of scales, highlighting the hierarchical nature of the cosmic web and allowing for better identification of the tenuous structures while still preserving the significant features of the cosmic web. Nonetheless, cluster regions signified by the white and yellow environments, are identified in this slice as filaments by NEXUS+ which only reiterates the fact that due to poor resolution, the density disperses into the surrounding such that the cluster morphological signal isn't as dominant. Since filaments are the next densest environment, these cluster regions are identified as filaments. This merging of the two environments accentuates this idea that it isn't suitable to identify the different constituents solely based on their density properties, and a more mathematically rigorous approach needs to be adopted when delineating the cosmic web.

In addition to the tracing of ascending manifolds, by being a topological based algorithm, DisPerSE is unable to characterise the thickness profile of filaments since it only extracts the skeleton (see Appendix F), and this is reflected by its lack of variety of different sized filaments detected. The lack of a thickness profile causes the detected filaments to be thicker than other formalisms including NEXUS+ (Libeskind et al., 2018) making it absorb both void-like and wall-like regions, giving for further false-detections as well as a loss of information on the fragile features. Contrariwise, NEXUS+ can preserve the shape and size of the different environments by adopting a geometrical and scale-space approach. By using geometrical information encoded within the Hessian density matrix, NEXUS+ allows itself to obtain information on the size and shape of the different environments, and upon analysing it over a range of smoothing scales, NEXUS+ better filters the identified environment giving researchers a nicely user-free rendered environment. Subsequently, NEXUS+ provides more reliable quantitative results.

Figure 5-3 shows the similarities and differences in filamentary detections between both formalisms. Starting with the similarities, we see that the figure is very similar to the one observed by NEXUS+ in figure 5-2, suggesting that NEXUS+ either under detects filaments or provides fewer false detections. In general, we can see that both formalisms allow tracing of the significant structures as the contours overlap the denser, brighter regions and traces well the network seen.

On the other hand, the figure on the right-hand side has its filamentary detections on the outskirts of the significant filaments, as well as a few detections situated in tenuous regions. DisPerSE has much more of these non-common detections due to the thickness of the network it portrays. More interestingly is that NEXUS+ has smaller length uncommon detections as well as a smaller fraction of these detections being in sparse regions. This difference is noteworthy as it suggests that DisPerSE not only removes the tenuous features that branch off significant filaments, but also reiterates the fact that DisPerSE has many false detections by not only identifying filaments as regions connected by density maximums, but also needing a user-defined thickness giving rise to systematic errors in its rendering of the cosmic web.

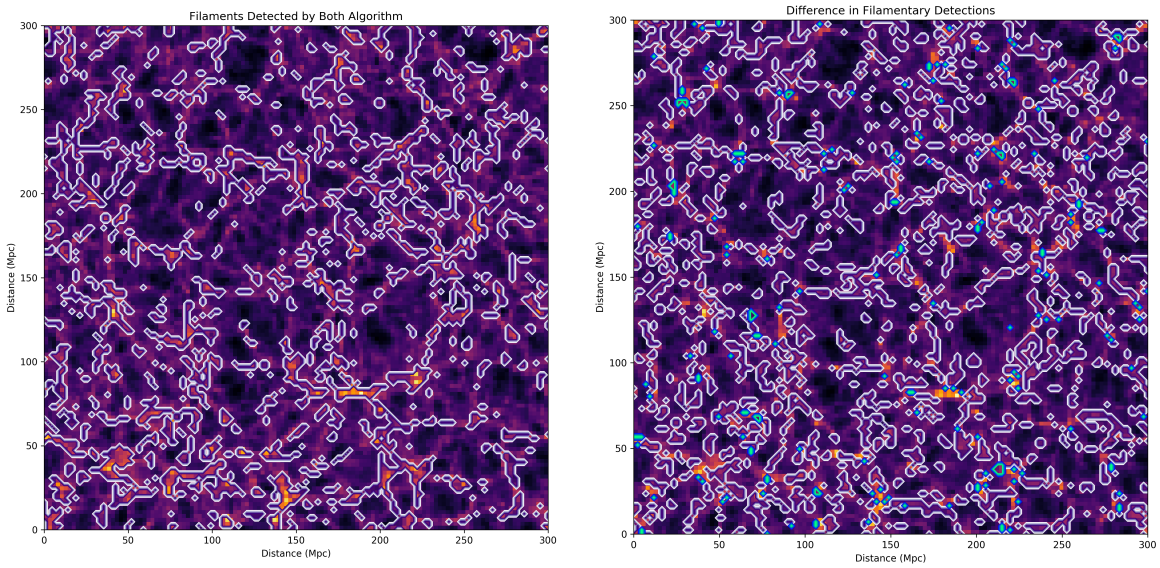


Figure 5-3: Left: Figure showing the common filaments detected between both formalisms. Right: Figure showing the filaments detected only by either NEXUS+ (highlighted in light blue) or DisPerSE (highlighted in purple). The slice for the detection's for this figure are projected along an infinitesimal slab at $z = 47$ Mpc.

When comparing by eye, one concludes that DisPerSE tends to over-identify filaments whereas NEXUS+ tends to under detect them. It is harder to make a conclusion on the tracing NEXUS+ provides since the figure shows a slice of the given Universe while the network we observe in the underlying density field can go much deeper than the observed slice, possibly explaining this lack of detection. Nonetheless, based on the differences, one concludes that although persistence provides a mathematically rigorous tool in the filtration of the cosmic web, it also removes some of the more intricate structures making up the boundary of filamentary networks and introduces systematic errors since it needs a user-defined scale when rendering the environment.

5.1.3 Voids

Figure 5-4 shows the voids identified by both formalisms. Since the figures show the identified voids with a 9 Mpc range in thickness, there may be identified regions that overlap with the filamentary network. Appendix I shows the void detections for an infinitesimal slice of the cosmic web as a comparison and helps emphasise the differences in both formalisms. The dots observed in both figures are due to the same projection effects discussed previously for filaments.

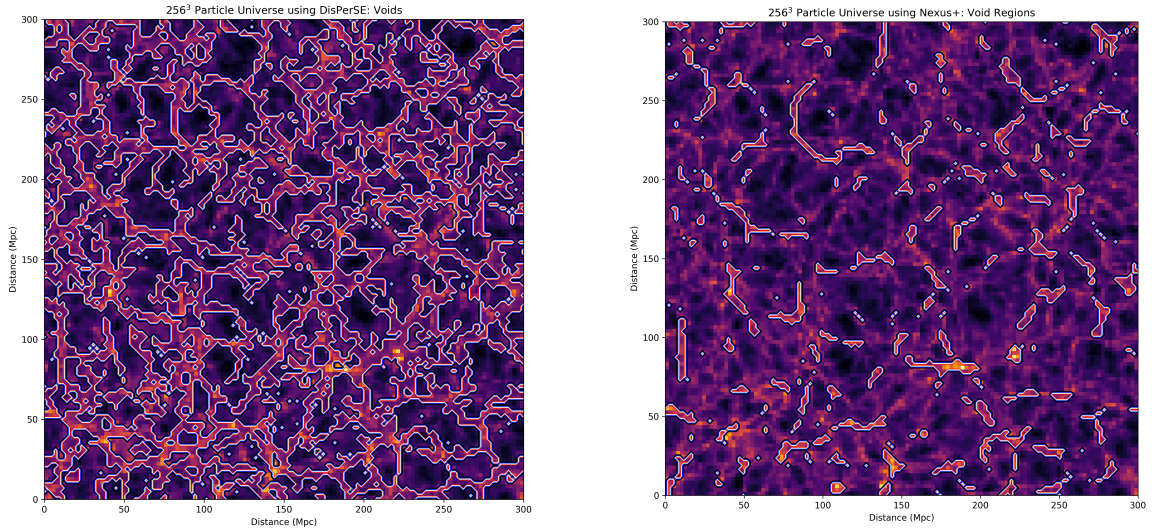


Figure 5-4: Two figures showing regions identified as voids for both DisPerSE (left) and NEXUS+ (right).

NEXUS+ identifies more voids compared to DisPerSE (see Appendix J for detections on the infinitesimal slice) and this is due to many factors, including the lack of a width profile DisPerSE obtains in its approach. For DisPerSE, we see a disconnected network of voids, with many being individual constituents appearing as patches. In contrast, NEXUS+ has the Universe engulfed by void regions, one that better correlates with the cosmic web theory and better traces the environment based on what we can delineate by eye.

Moreover, by construct NEXUS+ interprets all non-identified regions as voids, causing for an over-detection of them. Since NEXUS+ has a physically motivated threshold on the different morphological environments, underdeveloped and non-significant regions are often left undetected and end up identified as voids. This discrepancy in the number of detections between the two formalisms manifests itself in both the density distribution as well as volume and mass composition of the Universe as we will see in later sections of the report, however based on the rendering of the environment one can see that NEXUS+ provides better results when tracing the void regions of the cosmic web.

5.1.4 Walls

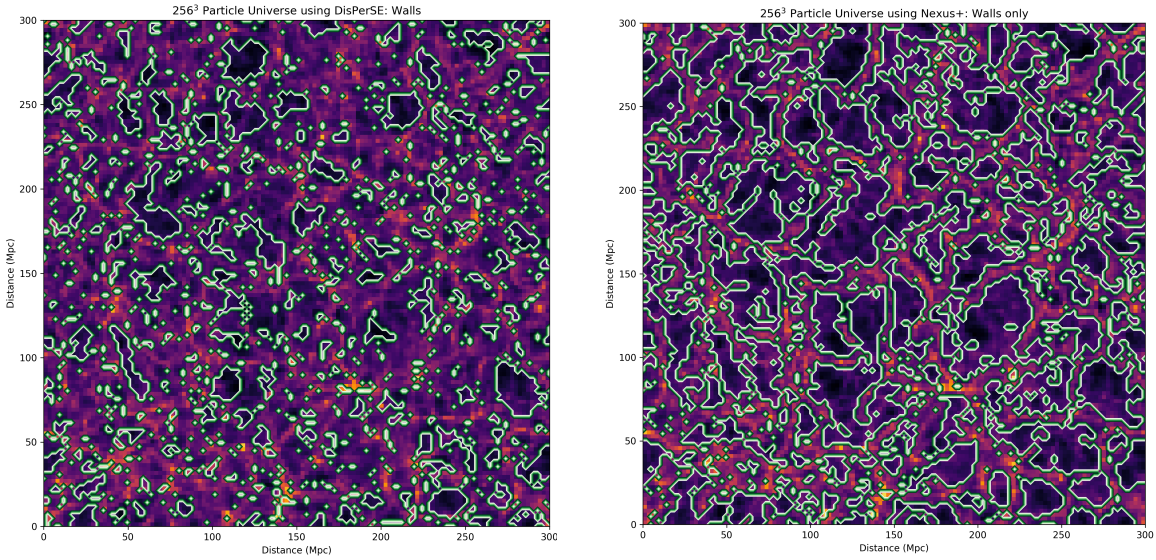


Figure 5-5: The detection of walls for the given Universe using both DisPerSE (left) and NEXUS+ (right).

Similarly to filaments, DisPerSE detects a larger amount of walls than NEXUS+. By analysing the cosmic web at a range of scales NEXUS+ analyses the multi-scale and hierarchical nature of the different environments, its log-Gaussian filter allows it to nicely delineate between void-like and wall-like regions which DisPerSE does not have the luxury in doing.

DisPerSE doesn't investigate the hierarchical nature of the cosmic web with the same success as NEXUS+. Its scale-free approach leads DisPerSE to ignore the concept of substructures and instead analyses every point as a specific structure without looking at the complete picture and how the different environments are connected. Furthermore, with walls being sparse and having a similar density profile to voids, a topological approach has difficulties when delineating between the two environments, often mistaking void regions as walls and vice-versa due to the lack of topological features within them.

Although identifying environments based on density is less than ideal due to the range of densities environments can consist of, a log-Gaussian filter allows NEXUS+ to identify regions with great success, including the more tenuous environments in walls.

Based on these qualitative observations, it is clear that both formalisms can trace the anisotropic pattern present in the cosmic web. Although the report uses an N-body simulation, this is also the case for observational data if we keep in mind that NEXUS+ is an improved version on MMF (Jones et al. (2010); Luber et al. (2019)) meaning that they both successfully meet two of the requirements needed for a successful cosmic web tracing formalism. With that being said, by eye, it is also clear that NEXUS+ traces the

environments better with the help of its geometrical analysis, but it should be stated that NEXUS+ tends to focus on giving better-rendered results whereas DisPerSE mainly focuses on the singularity structure of filaments.

Nonetheless, if quantifiable information needs to be extracted, researchers need an adequate rendering of results such that the rendered grid can then be applied to the underlying density field to extract the plethora of information ingrained within the cosmic web. This means that although NEXUS+ uses a second-order approximation as well as information ingrained within the density field to identify environments, the use of the Hessian density matrix allows for a better rendering of the different environments in a user-free way that preserves the shape and structure of the environment, resulting in better quantifiable results.

5.2 Mass and Volume Content of the Universe

The table below shows the mean density of the different environments:

	Filaments	Walls	Voids
DisPerSE	2.80	0.563	0.278
NEXUS+	5.73	1.29	0.408

Table 3: The average density, $\rho/\bar{\rho}$, of the different environments detected.

Due to clusters being undetected by DisPerSE, their density components are often absorbed within the filamentary network, while ones identified by NEXUS+ are ignored. Even so, NEXUS+ obtains a larger density for filaments due to DisPerSE identifying regions such as voids and walls to filaments, causing a decrease in the overall mean density of filaments.

The difference in mean density computed by either formalism manifests itself in the volume and mass composition of the Universe. The figures below show the volume composition of the Universe obtained by both formalisms.

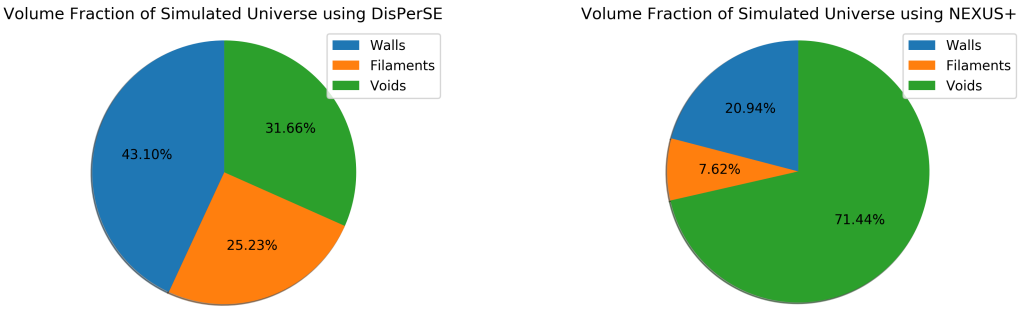


Figure 5-6: Volume fraction of the different constituents detected by either formalisms.

Since clusters make at most a few percent of the volume fraction of the Universe (Hahn

et al. (2007); Aragón-Calvo et al. (2010); Cautun et al. (2014); Libeskind et al. (2018)) its quantities will not alter the results exhibited here.

Based on previous results, one expected DisPerSE to have a more even distribution compared to NEXUS+ so this shouldn't come as a surprise. By not being able to detect the width of filaments and walls, both their volume fraction increase. This lack of preservation of shape or size of structures has such an effect that voids are no longer the dominant feature within the cosmic web, contrary to what the cosmic web theory expects. This stark contrast in the computed volume fraction of voids gives the most direct evidence that NEXUS+ provides a better analysis on the cosmic web, since it's own volume distribution better correlates with predictions of the cosmic web and better highlights the rich asymmetrical nature of the cosmic web.

Although NEXUS+ has the tendency to over-detect voids based on it identifying regions as the remaining environment, the increase it gets by absorbing under-developed structures doesn't induce a large difference in its computed volume fraction.

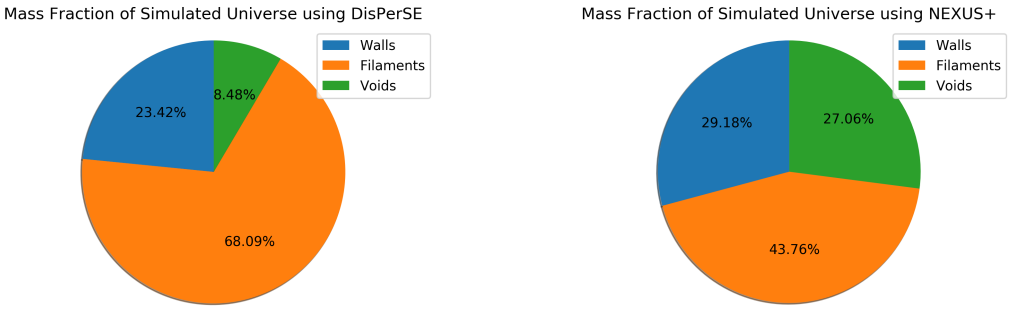


Figure 5-7: Mass fraction of the different constituents detected by either formalisms.

By over-detecting filaments, DisPerSE has a bias towards its mass composition, and with cluster regions having the possibility to be identified as filaments this is even more prominent a fact. Figure 5-7 illustrates this since filamentary detection make up an overwhelming part of the mass the Universe according to DisPerSE, while for NEXUS+, it is not as dominant a feature. Nonetheless, they both delineate filaments as the dominant mass constituent of the Universe coinciding with other algorithms as well as our real-life observations of the cosmic web since they are the most prominent constituent.

The approach NEXUS+ adopts in classifying voids also affects its computed mass fraction. By having underdeveloped regions filtered out, the mass comprising of these regions will be accounted for in the void fraction, causing both its mean density and mass fraction to increase.

The probability density function (PDF) in the figures below show that both formalisms

are successful in analysing the multi-scale range in densities environments can incorporate. DisPerSE has its trendlines skewed left as well as having narrower tails, signifying a smaller mean density as well as less variety in its results. This difference in distributions between the two formalisms is understood since filaments absorb both wall-like and void-like regions, who have low densities causing a shift towards lower values. Following the same logic, voids tend to be falsely identified as walls, causing a shift in the density distribution of walls towards lower densities.

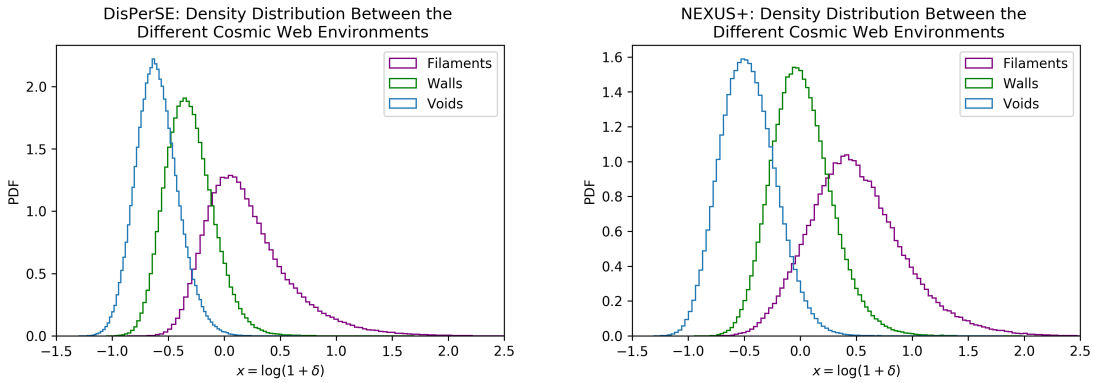


Figure 5-8: PDF of the different environments for both DisPerSE (left) and NEXUS+ (right).

The differences in the results found in the qualitative section of the analysis manifest itself in the quantitative results. Nonetheless, the results found shows that both formalisms successfully probe the multi-scale nature of the cosmic web in which environments can span a range of magnitudes. However, since one observes that the distribution function extracted by NEXUS+ has wider tails compared to those of DisPerSE it suggests that NEXUS+ provides a more coherent way in successfully analysing the multi-scale nature of the cosmic web.

5.3 Persistence vs. Scale-Space Filtration

Figure 5-9 shows the result of filamentary identification using different persistence and smoothing scale values.

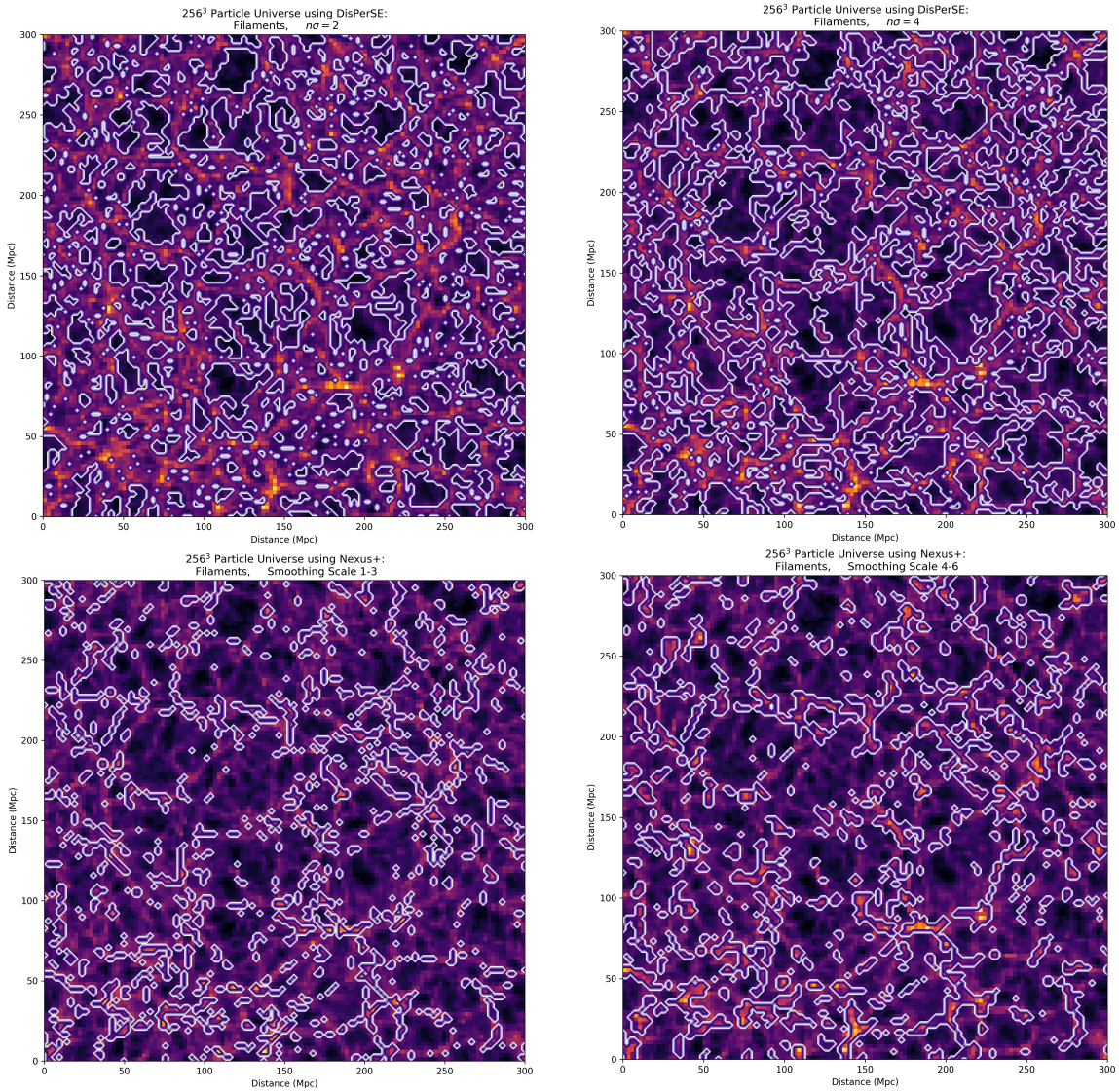


Figure 5-9: The detection of filaments for different persistent levels (top frame) and smoothing scales (bottom frame).

At first glance, we see that a $n\sigma = 2$ level practically covers the whole map only leaving the most underdense regions, signified by dark patches, undetected which is expected based on the discussion about figure 3-11. Within two significance levels, a more coherent network starts to form which traces the filaments nicely as void regions tend to empty first.

When looking at the difference in detection induced by different smoothing scales, there isn't as drastic a difference. Although the rendering varies, the lesser pronounced difference between smoothing scales not only implies that the geometric approach taken by NEXUS+ is reliable when incorporating a scale-space analysis, but also suggests that NEXUS+ is more adaptable to probing the hierarchical nature of the cosmic web.

Another way to look at the difference in detections both methods obtain and why an adaptable approach is essential for the analysis of the cosmic web, is by analysing the CDF plots seen in figure 5-10 situated on the next page.

A CDF plot emphasises the shift in the PDF distributions as it bins all identified points to its corresponding log-density ratio with respect to the overall cosmic mean density $\bar{\rho}$, and so one is better able to see the running total of a particular environments population. The three plots for DisPerSE all follow the same trend, $n\sigma = 2$ identifies more features with a lower density while $n\sigma = 4$ detects mostly dense, significant structures. The discussion provided in section 3.3.3 and figure 3-8 alluded to this fact since by increasing the persistence threshold a large fraction of tenuous features will get removed since their density contrast with the background isn't large, therefore being labelled as topologically insignificant. Trimming the tenuous features causes a lot of information on the intricacies of the cosmic web to be lost. Based on this, one concludes that although persistence is a mathematically rigorous method with lots of advantages, it may not be suitable for the analysis of the cosmic web - at least not in the fashion DisPerSE adopts that causes a local filtration of all structures based on a global criterion. This use of persistent homology to simplify the field could be improved upon in the algorithm by filtering features based on a local criterion. A local criterion would allow for the preservation of tenuous features but will also increase the computation time and depending on the fashion it is implemented, the number of user-parameters.

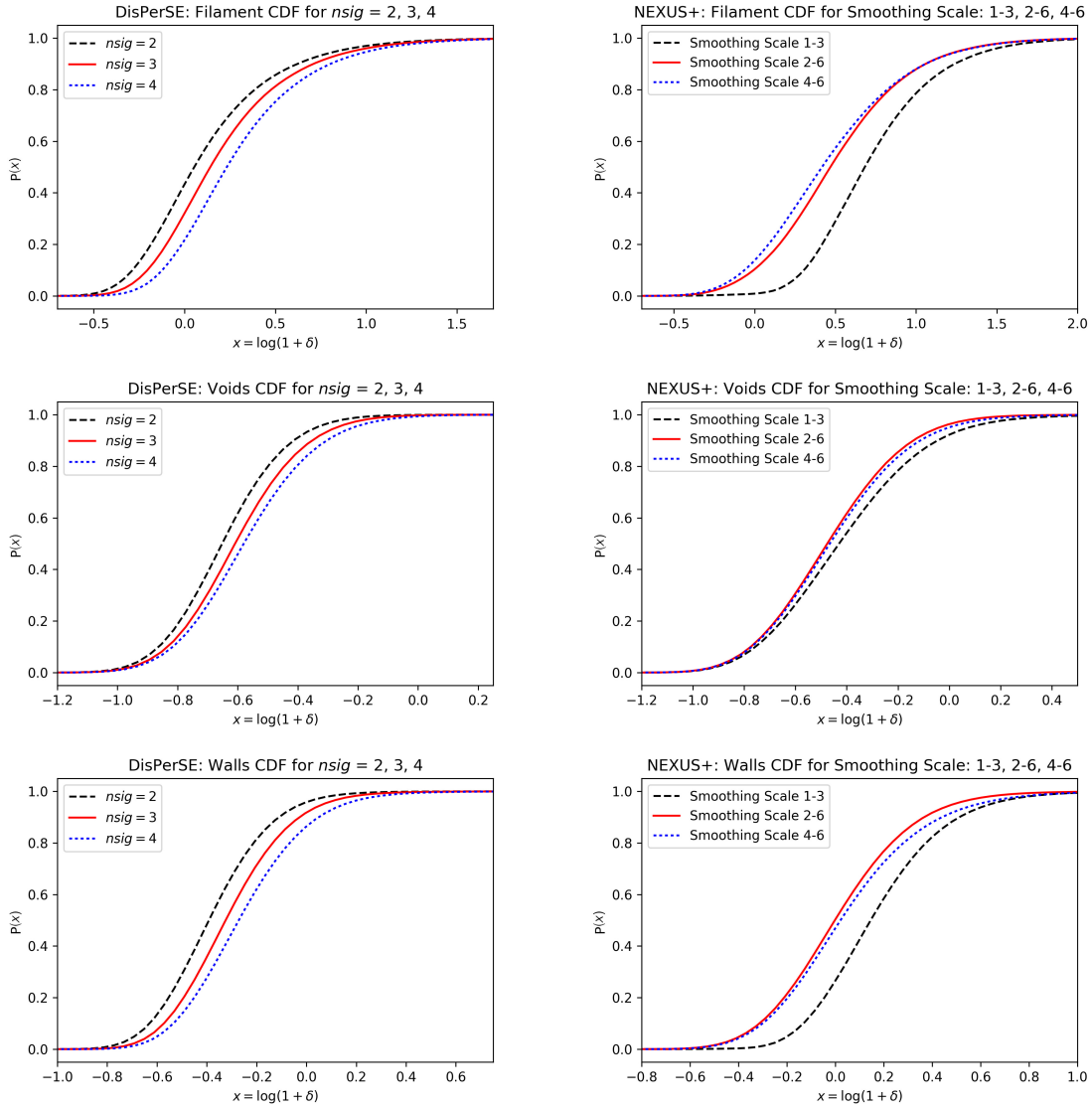


Figure 5-10: The CDF for the different constituents of the cosmic web. The left hand side shows the CDF based on DisPerSE detections and the right hand side the CDF based on NEXUS+ detections.

The CDF plots corresponding to NEXUS+ are more varied, showing its adaptable nature when analysing the environment. This flexibility is illustrated with the smoothing scale $4-6$ detecting more tenuous filaments than the other two filters while also having a larger fraction of its void and wall populations towards the higher densities compared to the smoothing scale $2 \leq n \leq 6$. The reason behind this is understood when looking at table 3, as well as the PDF's provided in figure 5-8. Since filaments tend to be the densest environment of the three, by observing only larger scales any region corresponding to a relatively large density will give a morphological signal related to filaments. Large smoothing scales analyse

large regions, and so dense, compact regions will be diluted with the surrounding void or wall-like regions inevitably causing the overall filament density to be reduced. This trend isn't reciprocated in the distribution exhibited for the other two environments since walls and voids have similar density values. Since the volume fractions of walls and voids make up a significant portion of the Universe, a large scale filter won't affect results as much when diluting the wall densities over a larger region, especially since it may also absorb filaments in the process which inevitably increases its density distribution.

Due to smoothing filter $1 - 3$ analysing at a smaller scale, noise becomes more dominant a feature and explains its bias towards denser structures. This trend hints towards some features detected at this scale being dependent on Poisson noise, giving support towards a scale-space approach since it is hard to define an optimal scale at which to analyse the cosmic web.

Analogous to the formalism itself, the difference between both filtration method is clear. One takes a mathematically robust approach in identifying topologically significant structures and provides a rigorous mathematical foundation towards detections, with the limitation being the removal of fragile features. The other formalism incorporates a scale-space mechanism in its filtration, allowing to investigate the intricate nature of the cosmic web and making it better in identifying tenuous structures and therefore, the connectivity between environments. Though its filtration method isn't as rigorous as the one adopted by DisPerSE, the scale-space approach taken by NEXUS+ is more adaptable, resulting in better identification of the substructures present as well as a better rendering of the environment.

6 Conclusion

The report begins by introducing the cosmic web and different ways researchers investigate its fragile nature. By building up knowledge on the mathematics behind both formalisms, section 5 provided a comparison based on results obtained using a pre-defined Universe. Although both algorithms were successful in tracing the different environments of the cosmic web and successful in showing the multi-scale nature of the cosmic web, there was a large discrepancy in both qualitative and quantitative results.

Both formalisms account for the hierarchical and multi-scale nature of the cosmic web with the use of DTFE, which interpolates and reconstructs the density field based on the geometry of the particle distribution in a scale-free, adaptable manner. Nonetheless, they do so with varying success. By adopting a global persistence ratio on a local neighbourhood, DisPerSE robustly trims out intricate features of the cosmic web whereas NEXUS+ with a scale-space approach is more adaptable, providing better results on both tenuous and significant features of the cosmic web, a fact highlighted with its wider tails in figure 5-8.

Furthermore, although DisPerSE provides a mathematical equivalence between the constituents with theorems, it is unable to characterise the shape or size of the different environments causing a poor rendering of results as well as introducing a systematic error as it needs a user-defined thickness. On the other hand, by identifying which cells correspond to which environment and analysing the geometrical information encoded in the second-order derivatives of the density field, NEXUS+ provides researchers with a user-free rendering on the environment by preserving both the shape and size of the different constituents and therefore yields better quantitative results as well as qualitative results.

It is clear upon looking at the two PDF's given in figure 5-8 that environments shouldn't be defined based solely on densities since they overlap. NEXUS+ avoids this by looking at many scales and only uses the information ingrained in the density field to identify regions possibly correlating to certain environments based on the best-fit result found for a wide range of scales. The strength in its rendering of the cosmic web is thanks to the geometrical information it obtains from the Hessian density matrix.

Based on these facts as well as the bonus of being less computationally intensive compared to DisPerSE, one concludes that the geometric, scale-space approach adopted by NEXUS+ is better at tracing and analysing the cosmic web compared to the topological approach implemented by DisPerSE.

Nonetheless, there is always room for improvement and the same can be said about NEXUS+. Though the method NEXUS+ adopts allows it to be successful in tracing the cosmic web, there are still improvements to be made. NEXUS+ doesn't provide a concrete mathematical equivalence to the different environments, and so ambiguity arises in its tracing of the cosmic web. Furthermore, the 4% discrepancy in results with respect to a continuous scale-space analysis may affect the identification of certain smaller features.

Future research may want to investigate how DisPerSE would perform on a better-sampled grid compared to NEXUS+ as the thickness of filaments between figures 3-8 and 5-2 already have a large discrepancy. Furthermore, the report focuses on a dark matter based simulation. Given more time, it would be insightful to test and compare either algorithm on observational data to understand how the formalism goes about fixing possible observational errors such as finger-of-god effects or magnitude as well as observational limitations, rendering lesser sampled results.

References

Robert J. Adler. *The Geometry of Random Fields: Classics in Applied Mathematics*. SIAM, Philadelphia, 1981.

M. A. Aragón-Calvo, B. J. T. Jones, R. van de Weygaert, and J. M. van der Hulst. The multiscale morphology filter: identifying and extracting spatial patterns in the galaxy distribution. , 474(1):315–338, October 2007. doi: 10.1051/0004-6361:20077880.

Miguel A. Aragón-Calvo, Rien van de Weygaert, and Bernard J. T. Jones. Multiscale phenomenology of the cosmic web. , 408(4):2163–2187, November 2010. doi: 10.1111/j.1365-2966.2010.17263.x.

J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. The Statistics of Peaks of Gaussian Random Fields. , 304:15, May 1986. doi: 10.1086/164143.

C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, E. Wollack, E. L. Wright, C. Barnes, M. R. Greason, R. S. Hill, E. Komatsu, M. R. Nolta, N. Odegard, H. V. Peiris, L. Verde, and J. L. Weiland. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results. , 148(1):1–27, September 2003. doi: 10.1086/377253.

J. Richard Bond, Lev Kofman, and Dmitry Pogosyan. How filaments of galaxies are woven into the cosmic web. , 380(6575):603–606, April 1996. doi: 10.1038/380603a0.

Nicholas A. Bond, Michael A. Strauss, and Renyue Cen. Crawling the cosmic network: identifying and quantifying filamentary structure. , 409(1):156–168, November 2010. doi: 10.1111/j.1365-2966.2010.17307.x.

V. Bonjean, N. Aghanim, M. Douspis, N. Malavasi, and H. Tanimura. Filament profiles from WISExSCOS galaxies as probes of the impact of environmental effects. *arXiv e-prints*, art. arXiv:1912.06559, December 2019.

T. Bonnaire, N. Aghanim, A. Decelle, and M. Douspis. T-ReX: a graph-based filament detection method. *arXiv e-prints*, art. arXiv:1912.00732, December 2019.

E. G. Patrick Bos. *Clusters, voids and reconstructions of the cosmic web*. PhD thesis, Kapteyn Astronomical Institute, December 2016.

Marius Cautun, Rien van de Weygaert, and Bernard J. T. Jones. NEXUS: tracing the cosmic web connection. , 429(2):1286–1308, February 2013. doi: 10.1093/mnras/sts416.

Marius Cautun, Rien van de Weygaert, Bernard J. T. Jones, and Carlos S. Frenk. Evolution of the cosmic web. , 441(4):2923–2973, July 2014. doi: 10.1093/mnras/stu768.

Marius Constantin Cautun. *The cosmic web and the local universe*. PhD thesis, University of Groningen, 2014. Relation: https://www.rug.nl/ Rights: University of Groningen.

Jörg M. Colberg, K. Simon Krughoff, and Andrew J. Connolly. Intercluster filaments in a ΛCDM Universe. , 359(1):272–282, May 2005. doi: 10.1111/j.1365-2966.2005.08897.x.

Peter Coles, Adrian L. Melott, and Sergei F. Shandarin. Testing approximations for non-linear gravitational clustering. , 260(4):765–776, February 1993. doi: 10.1093/mnras/260.4.765.

Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. volume 28, pages 454 – 463, 02 2000. ISBN 0-7695-0850-2. doi: 10.1109/SFCS.2000.892133.

Wiro J. Niessen Koen L. Vincken Max A. Viergever Frangi et al., Alejandro F. Multiscale vessel enhancement filtering. volume 1496, 02 1998. ISBN 978-3-540-65136-9.

M. Fukugita, C. J. Hogan, and P. J. E. Peebles. The Cosmic Baryon Budget. , 503(2):518–530, August 1998. doi: 10.1086/306025.

Punyakoti Ganeshaiah Veena, Marius Cautun, Rien van de Weygaert, Elmo Tempel, Bernard J. T. Jones, Steven Rieder, and Carlos S. Frenk. The Cosmic Ballet: spin and shape alignments of haloes in the cosmic web. , 481(1):414–438, November 2018. doi: 10.1093/mnras/sty2270.

Punyakoti Ganeshaiah Veena, Marius Cautun, Elmo Tempel, Rien van de Weygaert, and Carlos S. Frenk. The Cosmic Ballet II: spin alignment of galaxies and haloes with large-scale filaments in the EAGLE simulation. , 487(2):1607–1625, August 2019. doi: 10.1093/mnras/stz1343.

Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. The Geometry of Nonparametric Filament Estimation. *arXiv e-prints*, art. arXiv:1003.5536, March 2010.

Roberto E. González and Nelson D. Padilla. Automated detection of filaments in the large-scale structure of the Universe. , 407(3):1449–1463, September 2010. doi: 10.1111/j.1365-2966.2010.17015.x.

III Gott, J. R. The Sponge-Like Topology of Large Scale Structure in the Universe /. In Adelaide Hewitt, Geoffrey Burbidge, and Li Zhi Fang, editors, *Observational Cosmology*, volume 124 of *IAU Symposium*, page 433, January 1987.

III Gott, J. Richard, Mario Jurić, David Schlegel, Fiona Hoyle, Michael Vogeley, Max Tegmark, Neta Bahcall, and Jon Brinkmann. A Map of the Universe. , 624(2):463–484, May 2005. doi: 10.1086/428890.

J. E. Gunn, M. Carr, C. Rockosi, M. Sekiguchi, K. Berry, B. Elms, E. de Haas, Ž . Ivezić, G. Knapp, R. Lupton, G. Pauls, R. Simcoe, R. Hirsch, D. Sanford, S. Wang, D. York, F. Harris, J. Annis, L. Bartozek, W. Boroski, J. Bakken, M. Haldeman, S. Kent,

S. Holm, D. Holmgren, D. Petravick, A. Prosapio, R. Rechenmacher, M. Doi, M. Fukugita, K. Shimasaku, N. Okada, C. Hull, W. Siegmund, E. Mannery, M. Blouke, D. Heidtman, D. Schneider, R. Lucinio, and J. Brinkman. The Sloan Digital Sky Survey Photometric Camera. , 116(6):3040–3081, December 1998. doi: 10.1086/300645.

James E. Gunn and III Gott, J. Richard. On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution. , 176:1, August 1972. doi: 10.1086/151605.

Attila Gabor Gyulassy. *Combinatorial Construction of Morse-Smale Complexes*. PhD thesis, University of California, 2008.

Oliver Hahn, Cristiano Porciani, C. Marcella Carollo, and Avishai Dekel. Properties of dark matter haloes in clusters, filaments, sheets and voids. , 375(2):489–499, February 2007. doi: 10.1111/j.1365-2966.2006.11318.x.

Wojciech A. Hellwing. Dynamics of pairwise motions in the Cosmic Web. In R. van de Weygaert, S. Shandarin, E. Saar, and J. Einasto, editors, *The Zeldovich Universe: Genesis and Growth of the Cosmic Web*, volume 308 of *IAU Symposium*, pages 322–327, October 2016. doi: 10.1017/S1743921316010085.

Johan Hidding, Sergei F. Shandarin, and Rien van de Weygaert. The Zel'dovich approximation: key to understanding cosmic web complexity. , 437(4):3442–3472, February 2014. doi: 10.1093/mnras/stt2142.

Alex Ho, Max Gronke, Bridget Falck, and David F. Mota. Probing modified gravity in cosmic filaments. , 619:A122, November 2018. doi: 10.1051/0004-6361/201833899.

Bernard J. T. Jones, Rien van de Weygaert, and Miguel A. Aragón-Calvo. Fossil evidence for spin alignment of Sloan Digital Sky Survey galaxies in filaments. , 408(2):897–918, October 2010. doi: 10.1111/j.1365-2966.2010.17202.x.

Katarina Kraljic, Romeel Davé, and Christophe Pichon. And yet it flips: connecting galactic spin and the cosmic web. , 493(1):362–381, March 2020. doi: 10.1093/mnras/staa250.

Guilhem Lavaux and Benjamin D. Wandelt. Precision cosmology with voids: definition, methods, dynamics. *Monthly Notices of the Royal Astronomical Society*, 403(3):1392–1408, 04 2010a. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2010.16197.x. URL https://doi.org/10.1111/j.1365-2966.2010.16197.x.

Guilhem Lavaux and Benjamin D. Wandelt. Precision cosmology with voids: definition, methods, dynamics. *Monthly Notices of the Royal Astronomical Society*, 403(3):1392–1408, 04 2010b. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2010.16197.x. URL https://doi.org/10.1111/j.1365-2966.2010.16197.x.

Guilhem Lavaux and Benjamin D. Wandelt. Precision Cosmography with Stacked Voids. , 754(2):109, August 2012. doi: 10.1088/0004-637X/754/2/109.

Noam I. Libeskind, Rien van de Weygaert, Marius Cautun, Bridget Falck, Elmo Tempel, Tom Abel, Mehmet Alpaslan, Miguel A. Aragón-Calvo, Jaime E. Forero-Romero, Roberto Gonzalez, Stefan Gottlöber, Oliver Hahn, Wojciech A. Hellwing, Yehuda Hoffman, Bernard J. T. Jones, Francisco Kitaura, Alexander Knebe, Serena Manti, Mark Neyrinck, Sebastián E. Nuza, Nelson Padilla, Erwin Platen, Nesar Ramachandra, Aaron Robotham, Enn Saar, Sergei Shand arin, Matthias Steinmetz, Radu S. Stoica, Thierry Sousbie, and Gustavo Yepes. Tracing the cosmic web. , 473(1):1195–1217, January 2018. doi: 10.1093/mnras/stx1976.

Nicholas Luber, J. H. van Gorkom, Kelley M. Hess, D. J. Pisano, Ximena Fernandez, and Emmanuel Momjian. Large Scale Structure in CHILES. *arXiv e-prints*, art. arXiv:1904.10511, April 2019.

Yuichi Matsuda, Toru Yamada, Tomoki Hayashino, Hajime Tamura, Ryosuke Yamauchi, Takashi Murayama, Tohru Nagao, Kouji Ohta, Sadanori Okamura, Masami Ouchi, Kazuhiro Shimasaku, Yasuhiro Shioya, and Yoshiaki Taniguchi. Large-Scale Filamentary Structure around the Protocluster at Redshift z = 3.1. , 634(2):L125–L128, December 2005. doi: 10.1086/499071.

P. J. E. Peebles. *The large-scale structure of the universe*. 1980.

Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barand ela, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, E. Battaner, K. Benabed, A. Benoît, A. Benoit-Lévy, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. Bobin, J. J. Bock, A. Bonaldi, J. R. Bond, J. Borrill, F. R. Bouchet, M. Bridges, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, B. Cappellini, J. F. Cardoso, A. Catalano, A. Challinor, A. Chamballu, R. R. Chary, X. Chen, H. C. Chiang, L. Y. Chiang, P. R. Christensen, S. Church, D. L. Clements, S. Colombi, L. P. L. Colombo, F. Couchot, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, J. M. Delouis, F. X. Désert, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, J. Dunkley, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, T. C. Gaier, S. Galeotta, S. Galli, K. Ganga, M. Giard, G. Giardino, Y. Giraud-Héraud, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, J. Haissinski, J. Hamann, F. K. Hansen, D. Hanson, D. Harrison, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrand t, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, Z. Hou, W. Hovest, K. M. Huffenberger, A. H. Jaffe, T. R. Jaffe, J. Jewell, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, T. S. Kisner, R. Kneissl, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J. M. Lamarre, A. Lasenby, M. Lattanzi, R. J. Laureijs, C. R. Lawrence, S. Leach, J. P. Leahy, R. Leonardi, J. León-Tavares, J. Lesgourgues, A. Lewis, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, J. F. Macías-Pérez, B. Maffei, D. Maino, N. Mand olesi, M. Maris, D. J. Marshall,

P. G. Martin, E. Martínez-González, S. Masi, M. Massardi, S. Matarrese, F. Matthai, P. Mazzotta, P. R. Meinhold, A. Melchiorri, J. B. Melin, L. Mendes, E. Menegoni, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, C. B. Netterfield, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, I. J. O'Dwyer, S. Osborne, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, R. Paladini, D. Paoletti, B. Partridge, F. Pasian, G. Patanchon, D. Pearson, T. J. Pearson, H. V. Peiris, O. Perdereau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, P. Platania, E. Pointecouteau, G. Polenta, N. Ponthieu, L. Popa, T. Poutanen, G. W. Pratt, G. Prézeau, S. Prunet, J. L. Puget, J. P. Rachen, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, S. Ricciardi, T. Riller, I. Ristorcelli, G. Rocha, C. Rosset, G. Roudier, M. Rowan-Robinson, J. A. Rubiño-Martín, B. Rusholme, M. Sandri, D. Santos, M. Savelainen, G. Savini, D. Scott, M. D. Seiffert, E. P. S. Shellard, L. D. Spencer, J. L. Starck, V. Stolyarov, R. Stompor, R. Sudiwala, R. Sunyaev, F. Sureau, D. Sutton, A. S. Suur-Uski, J. F. Sygnet, J. A. Tauber, D. Tavagnacco, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, M. Tucci, J. Tuovinen, M. Türler, G. Umana, L. Valenziano, J. Valiviita, B. Van Tent, P. Vielva, F. Villa, N. Vittorio, L. A. Wade, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Wilkinson, D. Yvon, A. Zacchei, and A. Zonca. Planck 2013 results. XVI. Cosmological parameters. , 571:A16, November 2014. doi: 10.1051/0004-6361/201321591.

Adam G. Riess, Alexei V. Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M. Garnavich, Ron L. Gilliland, Craig J. Hogan, Saurabh Jha, Robert P. Kirshner, B. Leibundgut, M. M. Phillips, David Reiss, Brian P. Schmidt, Robert A. Schommer, R. Chris Smith, J. Spyromilio, Christopher Stubbs, Nicholas B. Suntzeff, and John Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. , 116(3):1009–1038, September 1998. doi: 10.1086/300499.

Johannes Rowland, Todd.Lipp and Eric W. Weisstein. Topological space. URL https://mathworld.wolfram.com/TopologicalSpace.html.

Nakajima S Shiraga N Atsumi H Yoshida S Koller T Gerig G Kikinis R. Sato et al., Y. Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. volume 2, 02 1998. ISBN 978-3-540-65136-9.

W. E. Schaap and R. van de Weygaert. Continuous fields and discrete samples: reconstruction through Delaunay tessellations. , 363:L29–L32, November 2000.

W. E. Schaap and R. van de Weygaert. *Delaunay recovery of cosmic density and velocity probes*, pages 483–484. 2003.

N. Schneider, T. Csengeri, M. Hennemann, F. Motte, P. Didelon, C. Federrath, S. Bontemps, J. Di Francesco, D. Arzoumanian, V. Minier, Ph. André, T. Hill, A. Zavagno, Q. Nguyen-Luong, M. Attard, J. Ph. Bernard, D. Elia, C. Fallscheer, M. Griffin, J. Kirk, R. Klessen, V. Könyves, P. Martin, A. Men'shchikov, P. Palmeirim, N. Peretto, M. Pestalozzi,

D. Russeil, S. Sadavoy, T. Sousbie, L. Testi, P. Tremblin, D. Ward-Thompson, and G. White. Cluster-formation in the Rosette molecular cloud at the junctions of filaments (Corrigendum). , 551:C1, March 2013. doi: 10.1051/0004-6361/201118566e.

T. Sousbie. The persistent cosmic web and its filamentary structure - I. Theory and implementation. , 414(1):350–383, June 2011. doi: 10.1111/j.1365-2966.2011.18394.x.

T. Sousbie, C. Pichon, S. Colombi, D. Novikov, and D. Pogosyan. The 3D skeleton: tracing the filamentary structure of the Universe. , 383(4):1655–1670, February 2008. doi: 10.1111/j.1365-2966.2007.12685.x.

White Simon D. M. Jenkins Adrian Frenk Carlos S. Yoshida Naoki Gao Liang Navarro Julio Thacker Robert Croton Darren Helly John Peacock John A. Cole Shaun Thomas Peter Couchman Hugh Evrard August Colberg Jörg Pearce Frazer Springel, Volker. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636, 03 2005. ISSN 1476-4687. doi: 10.1038/nature03597. URL https://doi.org/10.1038/nature03597.

R. Brent Tully. The Local Void is Really Empty. In Jonathan I. Davies and Michael J. Disney, editors, *Dark Galaxies and Lost Baryons*, volume 244 of *IAU Symposium*, pages 146–151, May 1987. doi: 10.1017/S1743921307013932.

R. van de Weygaert. The cosmic foam: stochastic geometry and spatial clustering across the universe. 6 2002.

R. van de Weygaert and J. Richard Bond. *Clusters and the Theory of the Cosmic Web*, volume 740. 2008. URL https://www.astro.rug.nl/~weygaert/tim1publication/weybondgh2005.paper1.pdf.

R. van de Weygaert and W. Schaap. Tessellation reconstruction techniques. 11 2000. doi: 10.1007/10849171_30.

R. van de Weygaert and W. Schaap. *The Cosmic Web: Geometric Analysis*, volume 665, pages 291–413. 2009. doi: 10.1007/978-3-540-44767-2_11.

Rien van de Weygaert, Miguel A. Aragon-Calvo, Bernard J. T. Jones, and Erwin Platen. Geometry and Morphology of the Cosmic Web: Analyzing Spatial Patterns in the Universe. *arXiv e-prints*, art. arXiv:0912.3448, December 2009.

Rien van de Weygaert, Gert Vegter, Erwin Platen, Bob Eldering, and Nico Kruithof. Alpha Shape Topology of the Cosmic Web. *arXiv e-prints*, art. arXiv:1006.2765, June 2010.

Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete and Computational Geometry*, 33:249–274, 02 2005. doi: 10.1007/s00454-004-1146-y.

Alfra J. Zomorodian. *Topology for Computing*. 2005.

Appendix A: Voronoi Tessellations

Voronoi tessellations have been heavily used in the research of the cosmic web and form the foundation of the Voronoi Clustering Model (van de Weygaert and Schaap, 2009). The Voronoi Clustering Model links the different morphological environments to the different components of a cell in an idealised manner as can be seen in table 4, providing a link between Voronoi tessellations with the natural pattern observed of the cosmic web induced by the dynamics of matter. Voronoi tessellations appear to naturally represent asymptotic configurations for a range of gravitational instability scenarios and also depicts the reason behind void regions being such a prominent feature within the cosmic web (van de Weygaert, 2002).

Geometric Component	Cosmic Web Equivalent
Voronoi Cell	Void
Voronoi Wall	Walls
Voronoi Edge	Filaments
Voronoi Vertex	Clusters

Table 4: The different geometric components and their corresponding morphological environments.

A Voronoi tessellation constitutes of a manifold partitioned into numerous cells that show the set of points closest to it's respective nuclei, which for DTFE is taken as the critical point. Mathematically this is written as (Zomorodian, 2005):

Definition A.1 (Voronoi Cell): A Voronoi cell V_u of $u \in \mathbb{R}^d$ is the set of points for which u minimises the weighted distance such that:

$$V_u = \{x \in \mathbb{R}^d | g_u(x_u, y) \leq g_v(x_v, y), \forall v \in K\}$$

and the union of Voronoi cells makes the Voronoi tessellation $V = \bigcup_i V_{u,i}$ where g_u is the Euclidean weighted distance between x and y. Figure A.1 illustrates the concept.

One can imagine the nuclei of each Voronoi cell being a peak in its local neighbourhood. By placing a ball at the maxima, a Voronoi cell shows the possible set of paths the ball can travel, the boundary showing the possible final destinations making it analogous to ascending and descending manifolds.

As one can expect from definition A.1, Voronoi tessellations are extremely sensitive to the geometrical composition of the discrete sample, and this is in part due to its dependence on the relation of clustering extent and clustering strength of the node. This sensitivity has allowed for reliable reconstruction of density fields, and its properties have allowed it to replicate the super-Hubble-like velocity flows found in voids (Schaap and van de Weygaert, 2003). The node distribution roots itself into the underlying field by being a topological invariant that isn't dependent on the galaxy population of the walls, filaments or cluster elements of the tessellation. Subsequently, this topological property allows for

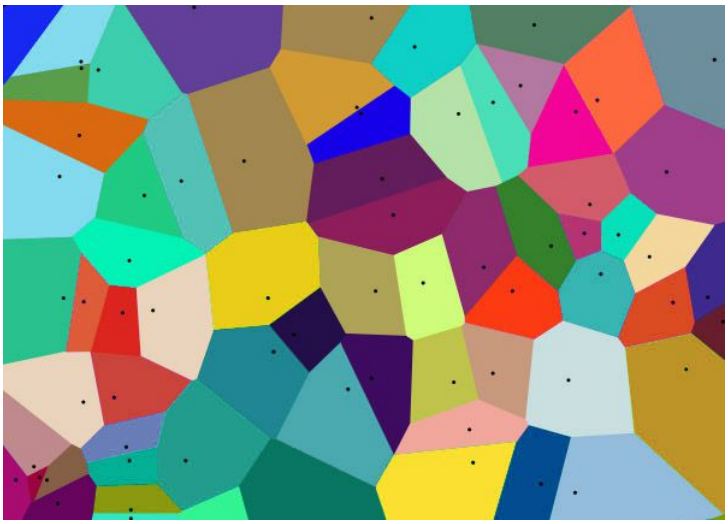


Figure A.1: Example of a weighted Voronoi 2D tessellation where the black dots show the nuclei of the cell.

robust quantitative measurements about the properties of the sampled matter distribution.

Delaunay tessellations are the dual of the Voronoi tessellation since one can construct a Delaunay tessellation by connecting the nuclei of the Voronoi cells, a mathematical property DTFE revolves around. Figure A.2 shows the relation between the two tessellations.

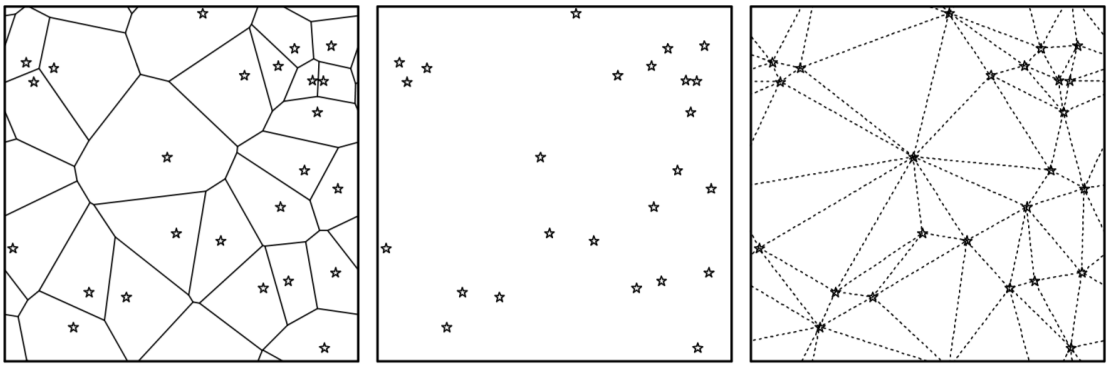


Figure A.2: The left frame shows the Voronoi tessellation of a given discrete sample set, shown in the central frame. The right-hand side shows its dual, the Delaunay tessellation. (image courtesy of van de Weygaert and Schaap (2000))

Appendix B: Simplices

Simplices are essential within computational analysis due to computers having a finite amount of storage, making it difficult for them to represent surfaces. Simplices allows a reduction of both the processing time and computational resources used by being the simplest d-dimensional geometric object in Euclidean space of dimensions d that preserves the topological and geometrical properties of a given sample (Zomorodian, 2005).

Both DisPerSE and NEXUS+ use DTFE, which takes advantage of different Delaunay tessellation properties to separate topological properties with geometrical properties of a given domain. It is therefore essential to provide a brief overview of some of the jargons within the mathematical formalism.

Definition B.1 (d-Simplex): A d-simplex, α_d, is a geometrical structure in a d-dimensional space which has $d+1$ points. This notion is illustrated below in figure B.1.

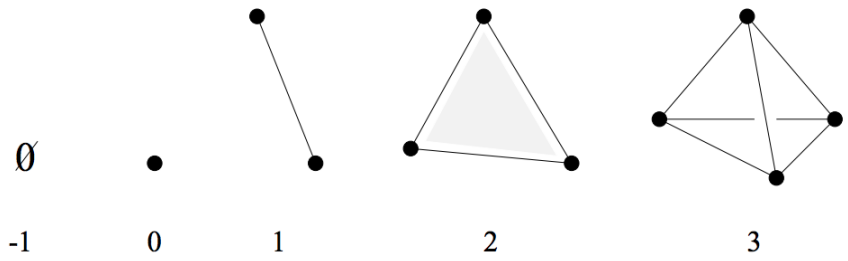


Figure B.1: Examples of simplices corresponding to different dimensions (numbered under).

Another way of interpreting a simplex is that a d-simplex is the convex hull of $d+1$ points, a convex hull being the smallest envelope to enclose all $d+1$ points within the set.

DisPerSE needs to compute values of the field at different regions to define it's discrete Morse function. It does this by giving a certain weight to each simplex, including its facets within the simplicial complex, but to understand the process better a few more definitions are provided. Most definitions are from Sousbie (2011), whereas the definition of a facet is from Gyulassy (2008).

Definition B.2 (Facet): Given an individual d-simplex, α_d, is an element of the simplicial complex K where $\dot{a}$ denotse it's boundary, the cells $\sigma_l \in K \cap \dot{a}$ denote the facets of α_d. If the difference in the dimension is one, $d-l=1$, we say that σ_l is the face of α_d.

Definition B.3 (Co-facet): For a simplex α_d within the simplicial complex K, a co-facet, β_l, is a simplex of a higher dimension, that is $l > d$. Here a parallel with the previous definition can be seen with α_d being a face of β_l if $l-d=1$.

A facet of a particular simplex α_d is a simplex which has a lower dimension than α_d, while

a co-facet is a simplex with a higher dimension than α_d. The following figure illustrates this idea:

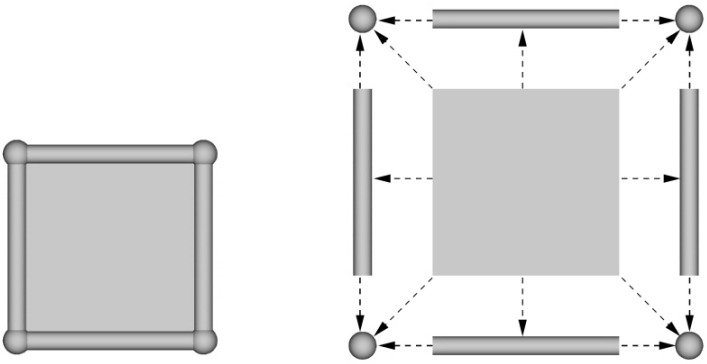


Figure B.2: Illustration conveying the idea of facets and co-facets. In this example, the face of the sheet-like surface would be the 1-simplicial edges while the 0-simplicial vertices are its facets. For the edges, the vertices are its face and the surface its co-face. (image courtesy of Gyulassy (2008))

Definition B.4 (Simplicial Complex): A simplicial complex K is a finite set of simplices such that obeys the following properties:

1. Any face of a simplex in K also belongs to K. This means that for $\alpha_d \in K$, σ_{d-1} also belongs to K.
2. Any intersection of two simplices in K is either empty or a simplex of a dimension lower or equal to the highest dimension they share.

A simplicial complex will partition space into a set of d-simplices which span the whole space based on the discrete points given acting as vertices. This explains why simplices are fundamental in computational analysis, as it allows to preserve the geometrical and topological properties of a surface while using the least resources to do so and spanning the whole d-dimensional Euclidean space.

Appendix C: Manifolds

DisPerSE revolves around manifolds and so an explicit definition is necessary to properly understand the analogies between the different cosmic web elements and d-dimensional ascending and descending manifolds. To understand what a manifold is, a few definitions are taken from Gyulassy (2008).

Definition C.1 (Set): A set is defined as a finite or infinite collection of elements a. $a \in S$ means that the element a is in the set S. On the other hand, if a set S contains no elements and is empty, it is called the null set and is represented by $\varnothing$.

Definition C.2 (Subset): A set U is a subset of set S if all the elements in U, $a \in U$, also make up the set S. This is denoted as $U \subseteq S$.

Definition C.3 (Union): The union of two sets, indicated with $\cup$, gives a new set consisting of all elements in A and B and is denoted as $A \cup B$. Figure C.1 illustrates this idea below.

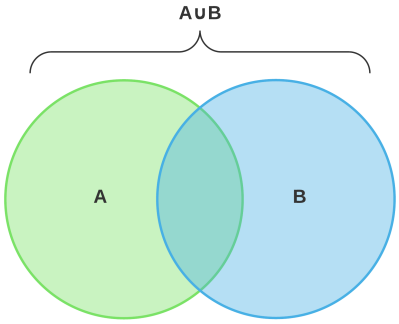


Figure C.1: The union between two sets A, B denoted as $A \cup B$.

Definition C.4 (Intersection): The intersection between two sets A and B is denoted as $A \cap B$. It is the set consisting of all the common elements between sets A and B and is figure C.2 represents this.

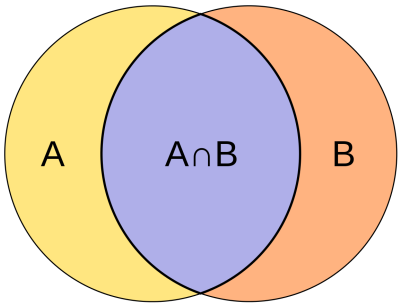


Figure C.2: The intersection between two sets A, B denoted as $A \cap B$ is highlighted in purple.

The following definition is taken from MathWorld (Rowland and Weisstein) and describes a topological space.

Definition C.5 (Topological Space): A topological space is a set S, along with a collection of open subsets U, that obey the following three axioms;

1. The null set $\varnothing$ as well as the set U belong to S. Mathematically: $\varnothing, U \subseteq S$.
2. The union of i sets in S, whether it is a finite or infinite amount, are also in S. Mathematically this is given as $\bigcup_i S_i \in S$.
3. The intersection of a finite or infinite amount of sets in S are also in S such that $\bigcap_i S_i \in S$.

Definition C.6 (Hausdorff space): A topological space X is called a Hausdorff space if for every $x, y \in X$ where $x \neq y$, there exists neighbourhoods A, B of x, y such that $A \cap B = \varnothing$. This means that two points x and y can be separated from a topological space if their neighbourhoods have no common elements between one another and figure C.3 helps illustrate this.

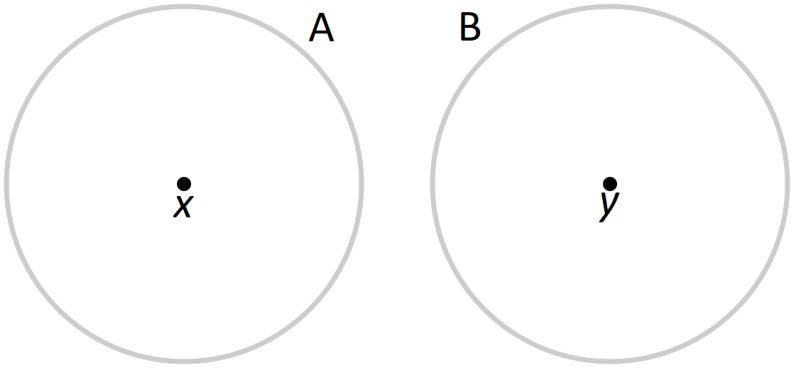


Figure C.3: An illustration of a Hausdorff space.

Given the definition of a homeomorphism given in section 3.3, the formal definition of a manifold can now be introduced.

Definition C.7 (Manifold): A separable Hausdorff space X, that is a space with a finite amount of neighbourhoods, is a d-manifold if at every point $x \in X$ it has a neighbourhood homeomorphic to $\mathbb{R}^d$.

Another way to approach its definition is that for every point $p = \{p_0, p_1 ... p_i\} \in S$, there is an open neighbourhood U of these points p as well as a homeomorphism $f : U \to V$ that maps the set U onto $V \in \mathbb{R}^d$.

This means that a d-dimensional manifold is a topological space in which its local neighbourhood looks like the d dimensional Euclidean space $\mathbb{R}^d$. Therefore a 0-manifold is

a topological space that locally resembles a 0-dimensional Euclidean space, a 1-manifold one that locally resembles a 1-dimensional Euclidean space etc.

In the given context a zero-dimensional manifold, 0-manifold, can be perceived as a vertex, a one dimensional manifold (1-manifold) a curve, a 2-manifold a surface and a 3-manifold a volume. This concept of manifolds is the fundamental reason as to why DisPerSE defines clusters as 0-manifolds, filaments 1-manifolds, walls 2-manifolds and voids 3-manifolds in a way that is mathematically rigorous.

Appendix D: Betti Numbers

Homology is a mathematical formalism that provides a quantitative description of the connectivity of space by analysing the existence of holes via the description of their boundaries.

The Euler characteristics and Betti numbers are two topological properties which allow researchers to obtain quantifiable information on a given manifold. The Euler characteristic, χ, shows the alternating sum of simplices on the surface of an object and is given by:

$$\chi = V - E + F = \beta_0 - \beta_1 + \beta_2 ... + (-1)^d \beta_d \tag{10}$$

Where V denotes vertices, E edges and F the faces. However, the Poincaré duality implies that $\chi = 0$ for the set of all 3-manifolds. The Poincaré duality combined with the fact that the Euler characteristic denotes the alternating sum of simplices - analogous to Betti numbers - suggests that Betti numbers provide more of a descriptive analysis on the topology of a given domain.

Though Betti numbers are only briefly mentioned in the report due to them being out of the scope of the study, it is still necessary to define them since they are a topological invariant which allows DisPerSE to analyse the environment so rigorously by looking at particular features' lifetimes.

Betti numbers are topological invariants that represent the nth homology group. Homology is a mathematical formalism that distinguishes holes on a manifold based on the connectivity between elements within the constrained space and describes the boundaries between these holes in a mathematical construct called cycles. For a d-dimensional manifold there will be d different types of holes ranging from $d \in [0, d-1]$. These holes correspond to the different Betti numbers and provide a rigorous method when analysing the cosmic web.

Formal definitions from Edelsbrunner et al. (2000) as well as Sousbie (2011) are provided below to build up the required jargon needed to understand Betti numbers better.

Definition D.1 (k-Chain): A k-chain is defined as the subset of all simplices in a complex K of dimension k. For a complex spanning $\mathbb{R}^3$, this would entail the set of vertices, segments, facets or tetrahedrons. Mathematically a k-chain can be defined as;

$$c_k = \sum_{i=1} n_i \alpha_i, \text{ where } n_i = \{0, 1\} \text{ and } \alpha_i \in K$$

Notice that n_i can be either one or zero, and therefore not all simplices within the complex make up the k-chain.

Definition D.2 (k-Chain Group): A k-chain group is the set of all its k-chains, that is:

$$C_k = c_k + c'_k = \sum_i (n_i + n'_i)\alpha_i$$

Definition D.3 (Boundary Operator): The boundary operator is denoted as ∂_k, and the boundary of an α_k simplex denotes the collection of it's $k-1$ dimensional faces.

An important property of the boundary operator is that the boundary of a boundary is the null set, that is $\partial_{k+1}\partial_k = \varnothing$. This states that a boundary is a cycle, and since cycles have no boundaries, it forms an empty set. The boundary operator allows to connect different chain groups of different dimensions in the following manner:

$$...C_{k+1} \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1}...$$

Definition D.4 (kth Cycle Group): Z_k indicates the kth cycle group and it is the kernel of the boundary operator ∂_k. Following the property of the boundary operator, when the cycle group c has the boundary applied to it, it becomes an empty set such that $\partial_k(c) = 0$. Mathematically this is defined as:

$$Z_k \equiv \ker \partial_k = \{c \in C_k | \partial_k(c) = 0\}$$

Z_k can be understood as the set of k-chains with an empty boundary.

Definition D.5 (kth Boundary Group): B_k is used to represent the kth boundary group. This is the image of the boundary operator and represents the collection of $(k-1)$-chains that are boundaries of the k-chains. Formally this is defined as being:

$$B_k \equiv \text{im } \partial_k = \{d \in C_{k-1} | \exists c \in C_k : d = \partial_k(c)\}$$

From definitions D.3, D.4 and D.5 it can be seen that $B_k \subseteq Z_k \subseteq C_k$ and this is illustrated in figure D.1 below.

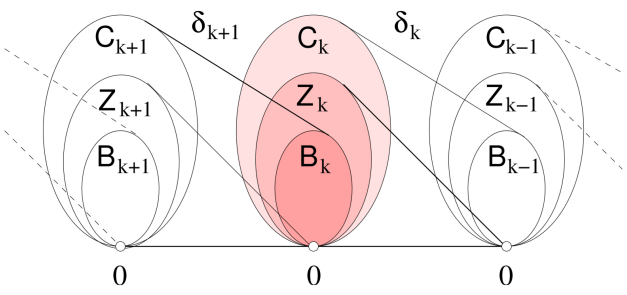


Figure D.1: Illustration showing the relation between chains, cycles and boundary groups with the boundary operator. (image courtesy of Edelsbrunner et al. (2000))

As can be seen, the kth boundary group makes up the $k+1$ chain group, while the kth cycle group has no boundaries in and of itself.

Definition D.6 (kth Homology Group): The kth homology group, H_k is given as $H_k = \frac{Z_k}{B_k}$. The kth homology group is the set of all the k-chains neglecting those of the boundary group. This gives rise to the possibility of separating groups based on their boundaries.

Definition D.7 (kth Betti Number): The kth Betti number, β_k, is the rank of it's kth homology group, or more formally $\beta_k = \text{rank } H_k$. It corresponds to the number of k-cycles in a given set.

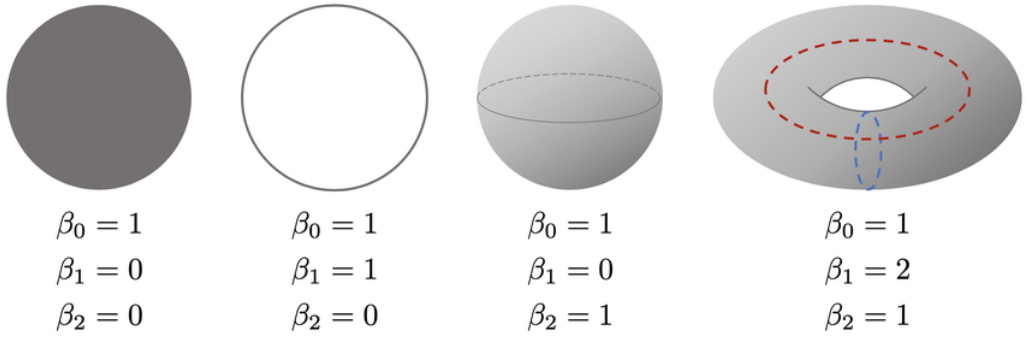


Figure D.2: Different surfaces along with the corresponding Betti numbers attributed to them.

For cosmology, there will be three different Betti numbers. Mentioned already in section 3.3, β_0 or the 0-cycle shows the number of isolated components of a surface, β_1 the number of tunnels and β_2 the number of enclosed volumes.

Figure D.2 gives examples of different surfaces and their Betti numbers. A flat circle that's filled in will have one independent component (the circle itself), as well as having no tunnels or holes. This gives rise to Betti numbers $\beta_0 = 1, \beta_1 = 0, \beta_2 = 0$ respectively. On the other hand, by observing the torus on the right-hand side of figure D.2 we see that the torus encloses a volume giving it a value of $\beta_2 = 1$ while also enclosing two tunnels, that is to say, it is possible to draw two different circles on the surface shown by the dashed red and blue lines amounting to a 1-cycle value of $\beta_1 = 2$. The concept of Betti numbers is important due to it providing a complete depiction of the topology of a given manifold, allowing for robust analysis of the field.

Appendix E: Pie Charts showing how the mass and volume composition of a 512^3 Universe depends on the smoothing scale.

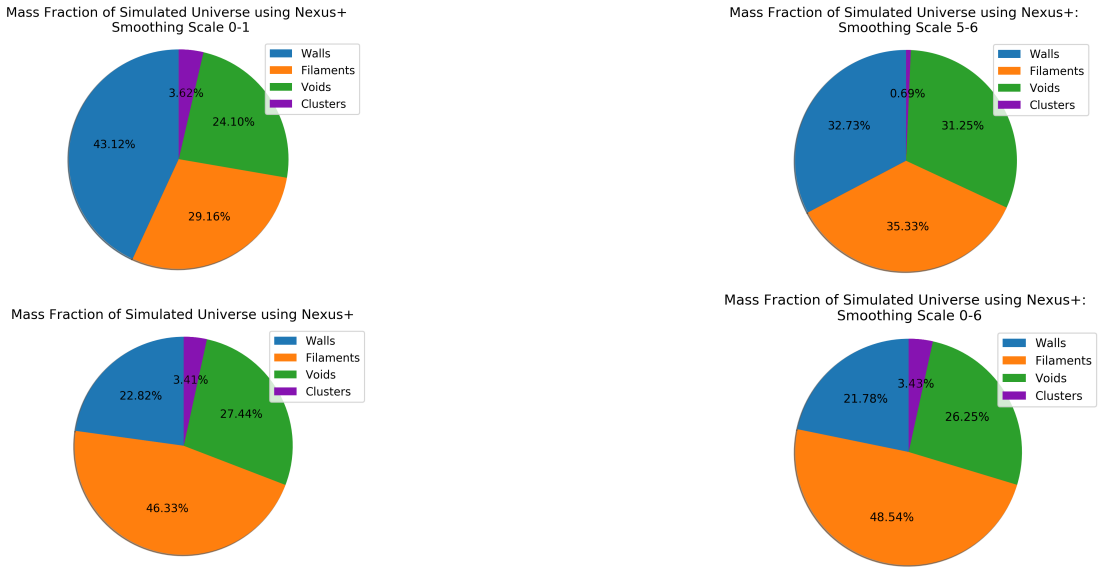


Figure E.1: Mass fraction of the Universe based on the smoothing scales used in the MMF. The unlabelled filter corresponds to the scale $2 \leq n \leq 6$ and is unlabelled due to being the default Universe upon the comparison with DisPerSE.

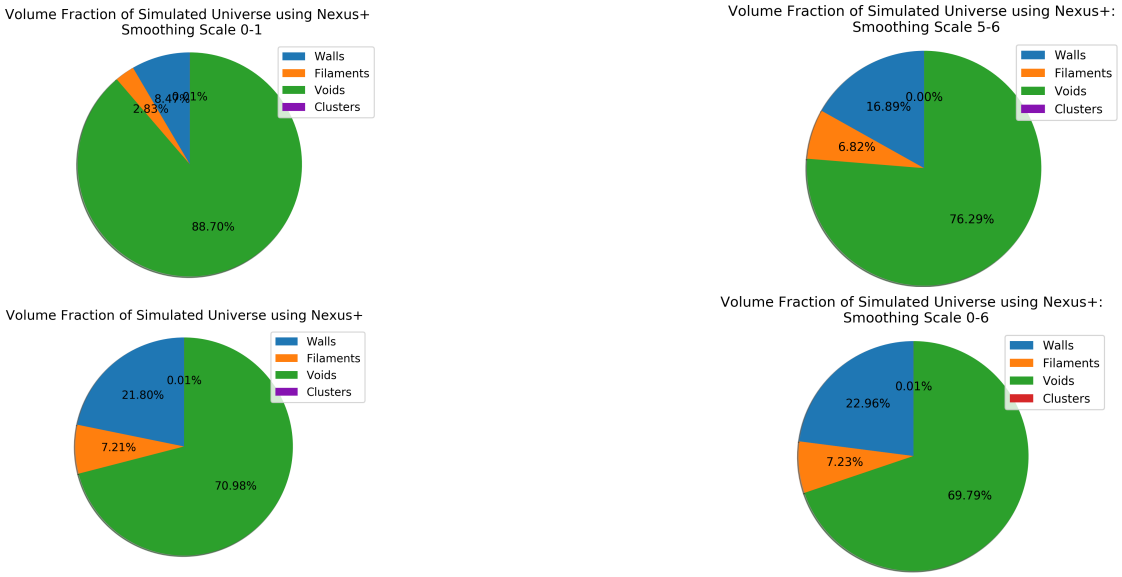


Figure E.2: Volume composition of the simulated Universe at a range of scales. Again, the unlabelled pie chart corresponds to the smoothing scale $2 \leq n \leq 6$.

Appendix F: Extracting the cosmic web's skeleton

This section provides a general overview of how DisPerSE characterises the skeleton of the filaments, for a more thorough discussion see Sousbie et al. (2008).

Due to the discrete nature of the data, Poisson noise gets introduced and potentially affects the shape of the different environments found by DisPerSE. DisPerSE computes a probability distribution dependent on the gradient flow to fix this error. By mapping the probability distribution on the field, pixels are processed by an algorithm to find which set of integral lines or environment it most likely corresponds to.

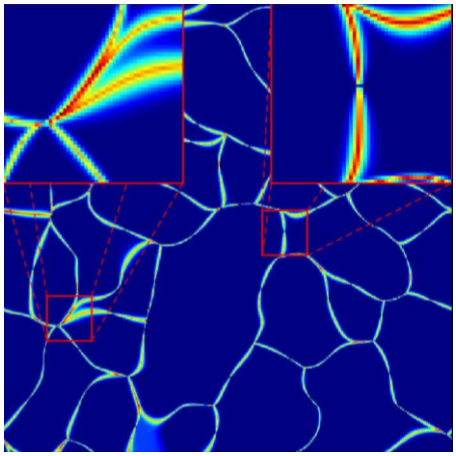


Figure F.1: Image showing the probability map that certain pixels are attributed to a skeleton. (image courtesy of Sousbie et al. (2008))

The algorithm computes a probability for each pixel based on the gradient flow and the corresponding probability distribution of its $3^d - 1$ neighbours, where d denotes the dimension of the data set provided. Figure F.1 illustrates this by showing the probability distribution of a filamentary network where red regions show areas most likely to be attributed as filaments. To construct the probability distribution the algorithm starts at the minimum (or void region) then iterates in an ascending fashion until a new local minimum is detected. At this point, it looks for the saddle point between the two critical points as this corresponds to regions where filaments will arise, and with the help of the probability distribution of each pixel, DisPerSE is then able to delineate the filament.

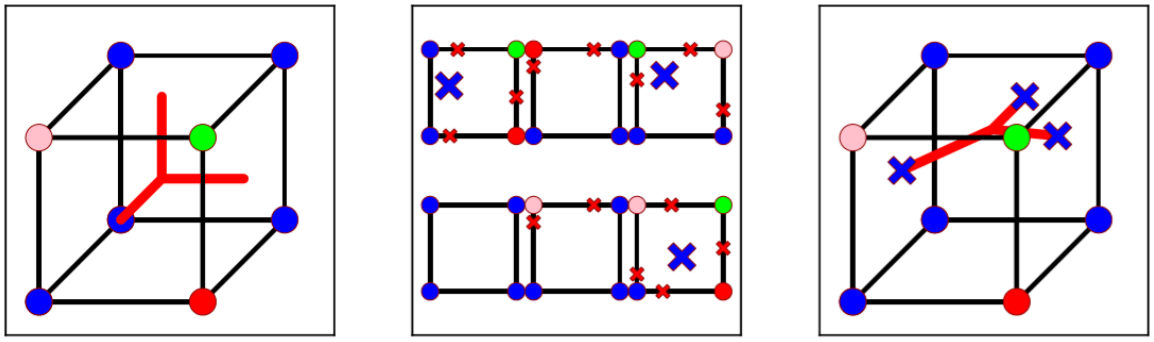


Figure F.2: The figure above illustrates the sub-pixel resolution delineation of filaments DisPerSE implements. Red crosses show the weighted average position where there is an equal probability that it belongs to one of the two pixels with different indices (blue denotes minima's, green saddle points and red maxima's). A blue cross then marks a new weighted average this time corresponding to a face of the cube based on the position of three or more red crosses that occupy the segments or edges. From these blue crosses, the filaments are traced. (image courtesy of Sousbie et al. (2008))

This method, however, still has issues with the structure being discontinuous due to resolution effects and so a further smoothing parameter has to be given to render better results. By taking eight neighbouring pixels in the field (forming the vertices of a cube), DisPerSE identifies each region whether it is a minimum, maxima or saddle point. For two neighbouring vertices in a cube where manifolds intersect, if their corresponding index $i \neq j$, DisPerSE computes a gradient weighted average position in which the filament is most likely to emerge. This weighted average between pixels gives for sub-pixel resolution and therefore better tracing of the skeleton and is illustrated in figure F.2. This process allows tracing of the ridges induced by the gradient flow of the underlying field and therefore, allows extraction of sub-pixel resolution of the skeleton of the network.

Appendix G: 3D representation of the filamentary network detected by both DisPerSE and NEXUS+.

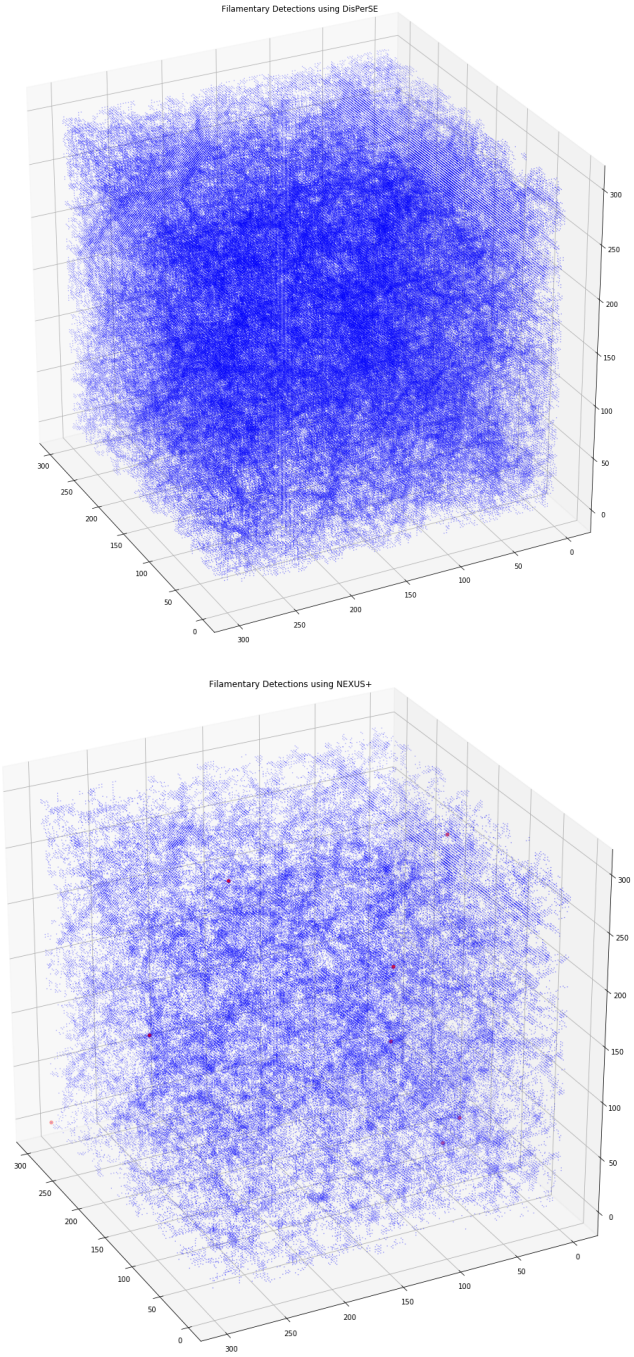


Figure G.1: Figure showing the identified filaments (blue dots) using both formalisms for the 256^3 particle Universe. Top frame shows the detections using DisPerSE and the bottom frame that of NEXUS+. The red in the NEXUS+ shows the detected nodes

Appendix H: Example of projection effects induced on detected filaments due to taking an infinitesimal slice.

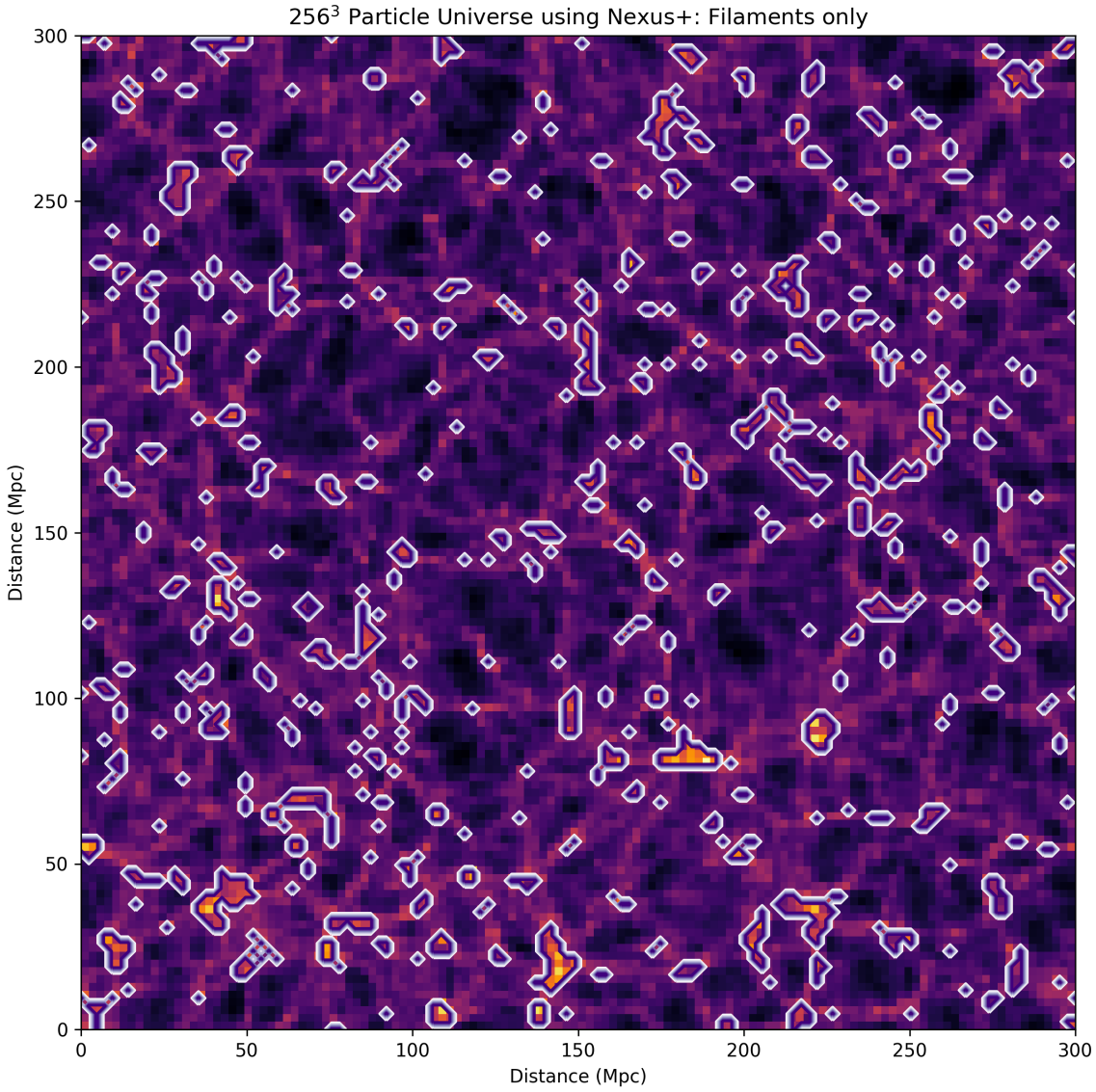


Figure H.1: Figure showing the projection effects acting on a filamentary network detected by NEXUS+. The detections are all points identified to be filaments at a $z = 47$ Mpc slice. The projection effect manifests itself in the identification of filaments by making a non-negligible fraction of the population appear as dots, this means that the filaments are not going sideways in the chosen slice and has a direction tangent to the plane we are observing.

Appendix I: Void regions identified in an infinitesimal slice of the simulated Universe.

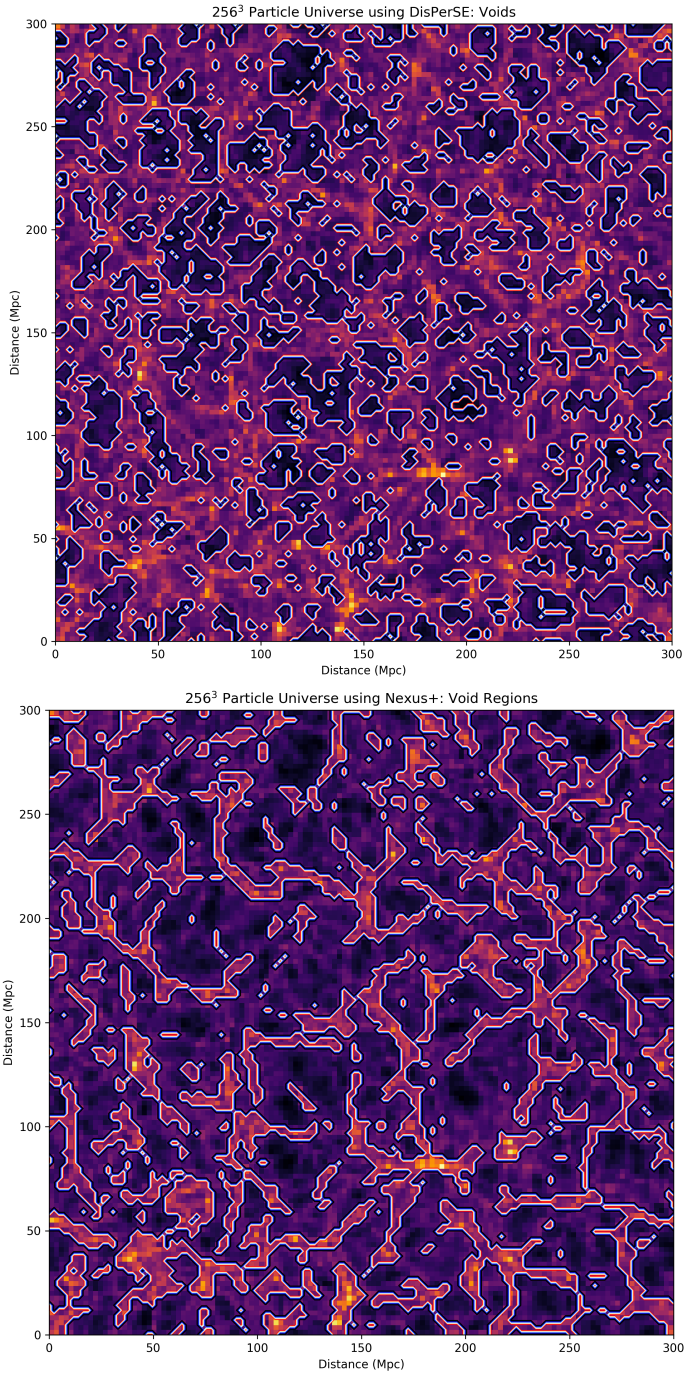


Figure I.1: Figure showing the identified void regions for both DisPerSE and NEXUS+ for an infinitesimal thickness. The two figures here show much better how the void regions form the boundary of the filamentary network and walls and the discrepancy between the results.

Appendix J: 3D representation of the voids detected by both DisPerSE and NEXUS+.

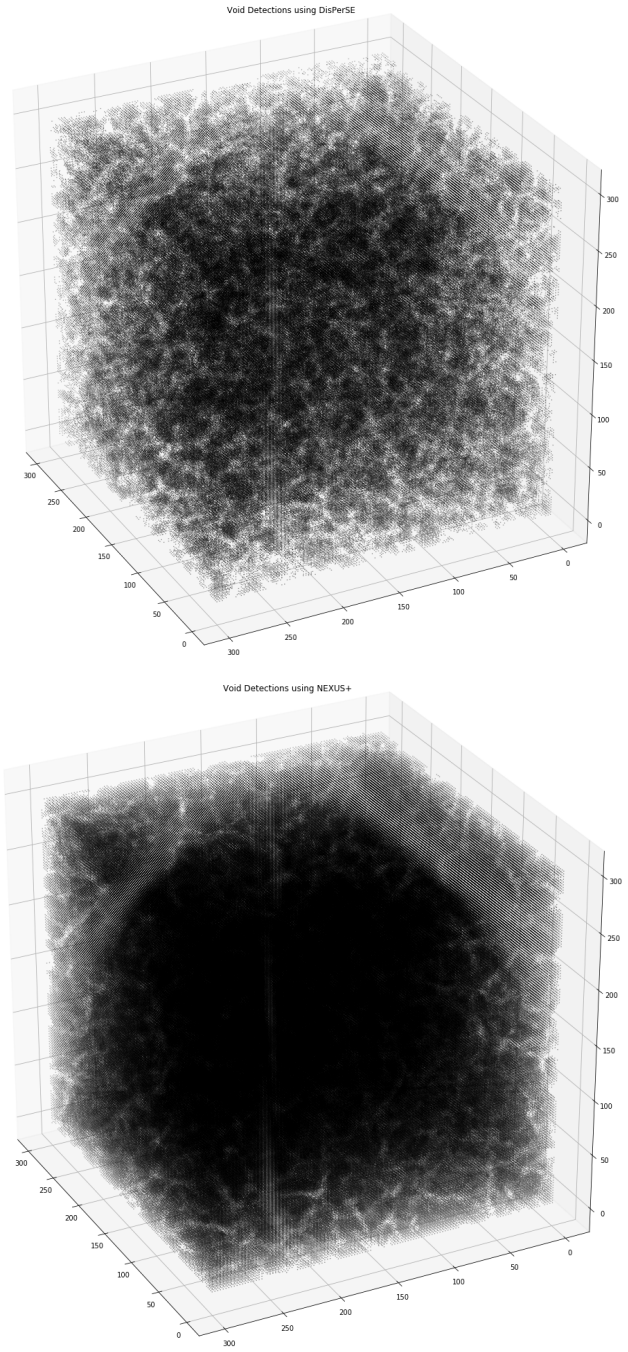


Figure J.1: Figure showing all the identified void regions using either DisPerSE or NEXUS+. The two figures further emphasise the discrepency in void identification between the two formalisms.

Appendix K: Zel'dovich Formalism

Though not necessary for our understanding of the two formalisms, the Zel'dovich approximation will be briefly discussed simply due to its significance on the study of the cosmic web as it provides a good description of the evolution of the cosmic web in the linear regime. The Zel'dovich approximation provides a natural way in describing the anisotropic collapse in the linear regime (as shown in Coles et al. (1993)) and has been used countless times as the initial conditions for N-body simulations.

Starting from the equation that traces the motion of an element;

$$x(t) = q_0 + D(t)\nabla\psi(q)$$

where q_0 is the initial position of a given element, $D(t)$ is the linear growth factor and $\psi(q)$ the Lagrangian displacement potential (for more information see Peebles (1980)), it is possible to rearrange it to obtain a function that describes how the mass of a particular element evolves as a function of time;

$$\rho(x) = \frac{\bar{\rho}}{(1 - D\lambda_1(q))(1 - D\lambda_2(q))(1 - D\lambda_3(q))}$$

where $\bar{\rho}$ is the average density of the environment and λ_i denotes the eigenvalue of the deformation tensor given as;

$$\psi_{ij}(q) = \frac{\partial^2\psi(q)}{\partial q_i \partial q_j}$$

This equation is imperative to the study of the cosmic web as it allows a classification of morphological environments based on the eigenvalues given in the deformation tensor ψ (see table 5), while also providing a natural description on the sequence of the formation of the constituents of the cosmic web as well as the anisotropic nature of their collapse.

The anisotropic nature of structure formation can be seen by the trend that in general $\lambda_1 \neq \lambda_2 \neq \lambda_3$ and the hierarchical nature of structure formation by the fact that the dominant direction of the collapse takes place towards the largest eigenvalue (by convention here, $\lambda_1 < \lambda_2 < \lambda_3$).

Structure	Eigenvalue Constraint
Cluster	$\lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0$
Filament	$\lambda_1 > 0, \lambda_2 > 0, \lambda_3 < 0$
Sheet	$\lambda_1 > 0, \lambda_2 < 0, \lambda_3 < 0$
Void	$\lambda_1 < 0, \lambda_2 < 0, \lambda_3 < 0$

Table 5: The different morphological environments and their corresponding eigenvalues.

As alluded to, the approximation breaks down once the simulation or the environment leaves the linear regime and different elements start to cross paths. This breakdown can be seen by a Taylor expansion of the density equation $\rho(x)$ and seeing how the density contrast relates with these eigenvalues (see Hidding et al. (2014) for more). More intuitively, once the paths of two different elements intersect the linear regime is no longer valid as their motion are now dominated by the gravitational fields of the non-linear structure (Cautun et al., 2014).