



**HAL**  
open science

# The persistent cosmic web and its filamentary structure - II. Illustrations

T. Sousbie, C. Pichon, H. Kawahara

► **To cite this version:**

T. Sousbie, C. Pichon, H. Kawahara. The persistent cosmic web and its filamentary structure - II. Illustrations. *Monthly Notices of the Royal Astronomical Society*, 2011, 414, pp.384-403. 10.1111/j.1365-2966.2011.18395.x . insu-03645953

**HAL Id: insu-03645953**

**<https://insu.hal.science/insu-03645953>**

Submitted on 22 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The persistent cosmic web and its filamentary structure – II. Illustrations

T. Sousbie,<sup>1,2\*</sup> C. Pichon<sup>2,3</sup> and H. Kawahara<sup>4</sup>

<sup>1</sup>Department of Physics, The University of Tokyo, Tokyo 113-0033, Japan

<sup>2</sup>Institut d'astrophysique de Paris & UPMC (UMR 7095), 98, bis boulevard Arago, 75 014 Paris, France

<sup>3</sup>Oxford Astrophysics, Department of Physics, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH

<sup>4</sup>Department of Physics, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan

Accepted 2011 January 19. Received 2011 January 14; in original form 2010 September 21

## ABSTRACT

The recently introduced discrete persistent structure extractor (DisPerSE, Sousbie, Paper I) is implemented on realistic 3D cosmological simulations and observed redshift catalogues; it is found that DisPerSE traces very well the observed filaments, walls and voids seen both in simulations and in observations. In either setting, filaments are shown to connect on to haloes, outskirt walls, which circumvent voids, as is topologically required by the Morse theory. Indeed this algorithm returns the optimal critical set while operating directly on the particles. DisPerSE, as illustrated here, assumes nothing about the geometry of the survey or its homogeneity, and yields a natural (topologically motivated) self-consistent criterion for selecting the significance level of the identified structures. It is shown that this extraction is possible even for very sparsely sampled point processes, as a function of the persistence ratio (a measure of the significance of topological connections between critical points). Hence, astrophysicists should be in a position to trace precisely the locus of filaments, walls and voids from such samples and assess the confidence of the post-processed sets as a function of this threshold, which can be expressed relative to the expected amplitude of shot noise. In a cosmic framework, this criterion is shown to level with the friends-of-friends structure finder for the identification of peaks, while it also identifies the connected filaments and walls, and quantitatively recovers the full set of topological invariants (number of holes, etc.) *directly from the particles*, and at no extra cost as a function of the persistence threshold. This criterion is found to be sufficient even if one particle out of two is noise, when the persistence ratio is set to  $3\sigma$  or more. The algorithm is also implemented on the SDSS catalogue and used to locate interesting configurations of the filamentary structure. In this context, we carried the identification of an ‘optically faint’ cluster at the intersection of filaments through the recent observation of its X-ray counterpart by *Suzaku*.

**Key words:** methods: data analysis – galaxies: formation – galaxies: kinematics and dynamics – cosmology: observations – dark matter – large-scale structure of Universe.

## 1 INTRODUCTION

Over the past decades, numerical simulations (e.g. Efstathiou et al. 1985) and large redshift surveys (e.g. de Lapparent, Geller & Huchra 1986) have highlighted the large-scale structure (hereinafter LSS) of our Universe, a cosmic web formed by voids, sheets, elongated filaments and clusters at their nodes (Pogosyan et al. 1996). Characterizing quantitatively these striking features of the observed and modelled universe has proven to be both useful (Sousbie et al. 2008a; Gay et al. 2010) and challenging. It is useful because these

features reflect the underlying dynamics of the structure formation and are therefore sensitive to the content of the universe (Pogosyan et al. 2009). It is challenging because observations and simulations provide limited and noisy data sets. Recently, Sousbie (2011, hereinafter Paper I) presented an algorithm able to identify the voids, walls, filaments and peaks (or technically speaking, its critical sets) from a given noisy discrete sample of an underlying field. Typically, this situation arises in astrophysics when the aim is to recover the topology or the geometry of the underlying density field, while only a catalogue of galaxies are available. For instance, in the context of understanding the history of our Milky Way, it is of interest to identify the filaments of the Local Group. Yet typically in this context, only a limited number of galaxies at somewhat poorly estimated

\*E-mail: tsousbie@gmail.com

positions are observed. For redshift catalogues involving hundreds of thousands of galaxies, one would also wish to reconstruct the main features of the cosmic web as perfect as the non-uniform sampling allows. From a theoretical point of view, it might, for instance, be of interest to compute the cosmic evolution of the filamentary network, as its history constrains the dark energy content of the universe. From an observational point of view, it could also help solving the missing baryon problem (Fukugita, Hogan & Peebles 1998) because most of such baryons have been considered to be located along the filamentary structure in the form of the diffuse hot gas called warm/hot intergalactic medium (WHIM, Cen & Ostriker 1999; Aracil et al. 2004). Identifying the filament from galaxy distributions clearly provides good candidates for searching the WHIM with ultraviolet absorptions (e.g. Tripp et al. 2000; Danforth et al. 2010), X-ray absorptions (e.g. Fang, Bryan & Canizares 2002; Kawahara et al. 2006; Buote et al. 2009; Fang et al. 2010) and X-ray emission lines by future surveys (e.g. Yoshikawa et al. 2003; Ohashi et al. 2006). It is therefore of prime importance to provide a tool which deals consistently with such possibly sparse discrete samples. Quite a few such options have been presented recently (Stoica et al. 2005; Novikov, Colombi & Doré 2006; Hahn et al. 2007; Platen, van de Weygaert & Jones 2007; Stoica, Martínez & Saar 2007, 2010; Neyrinck 2008; Sousbie et al. 2008a,b; Forero-Romero et al. 2009; Sousbie, Colombi & Pichon 2009; Aragón-Calvo et al. 2010a,b; Paper I), relying on different strategies on how to deal with these constraints (see also the nice review of Noh & Cohn 2011).

The companion paper, Paper I, presented an algorithm, DisPerSE, which recovers the important features of the underlying cosmic field even when little information is available, so that the procedure manages to identify the most-robust features of the field. Roughly speaking, DisPerSE extracts filaments as a set of connected segments, walls as sets of connected triangles and voids as sets of connected tetrahedrons. It also keeps track of their relative connectivity (e.g. it follows which walls a given set of filaments are bounding). The extraction can operate directly on the particles.<sup>1</sup> It requires only one tunable parameter which corresponds to the significance of the retained features in units of  $\sigma$ . *Topology* (in fact discrete topology) provided the context in which this algorithm was implemented in order to carry out the extraction. Topology de facto characterizes the ‘rubber’ geometry of the underlying field, that is, its most-intrinsic and robust features. More specifically, DisPerSE produces an ensemble of critical sets (lines, surfaces and volumes) consistent with those defined within the context of the Morse theory for sufficiently smooth fields (Milnor 1963; Jost 2008). Recall that the Morse theory basically provides a rigorous framework in which to formally define such sets for ‘regular’ density fields (here, regular basically means non-degenerate so that these sets are always well defined). For instance, the critical lines defined by this theory connect peaks and maxima via special (extremal) flow lines of the gradient. The Morse theory formalizes the process of partitioning space according to the gradient flow of the density into the so-called ascending and descending manifolds. In other words, it tags space according to where one would end up going ‘uphill’ or ‘downhill’. In doing so it identifies special lines or surfaces, where something unusual happens. These lines trace quite well the filaments of the LSS. Similarly, the walls of the LSS have a natural equivalent feature as the ‘critical’ surfaces of the Morse theory (the so-called 2-manifolds). Our implementation in DisPerSE proceeds within the context of its discrete counterpart (Forman 2002). This

discrete construction is as consistent as possible with all the topological features of the underlying smooth field (it globally preserves, at the level of the noise, the salient features of the field, such as the number of connected components, the number of tunnels or holes defined by its isocontours, etc.; conversely,<sup>2</sup> the significant discrete critical sets have the correct ‘combinatorial’ properties e.g. critical lines only connect at critical points and saddle points connect exactly two peaks together, etc.). The level of complexity of the corresponding network also reflects the inhomogeneities of the underlying survey, that is, it adapts its level of description to the sampling, hence yielding a parameter-free multiscale description of the cosmic web. In fact, it also provides a simple diagnostic in order to estimate the robustness of the various components of the network (i.e. the degree of reproduced details should be tunable to the purpose of the investigation). Finally, it clearly addresses the shortcomings of watershed-based methods, namely the occurrence of spurious boundary lines induced by resampling in 3D or more.

The two main novel features DisPerSE addressed were (i) defining the counterpart of the (topologically consistent) critical sets *on* the mesh; and (ii) defining a procedure to simplify the corresponding mesh at the level of the local shot noise. The first step was achieved by considering simultaneously all the discrete components of the triangulated mesh (its vertices, edges, faces and tetrahedra), and reassigning a density to all these components in a manner which is heuristically consistent with the sampled density at the vertices; this relabelling procedure also ensures that the discrete flow (which follows from the corresponding discrete gradient) is sufficiently well-behaved to provide such topological consistency (specifically, it ensures the existence of discrete counterparts of regular critical points). Amongst the discrete analogues of gradient flows, DisPerSE identifies the critical subsets as special sets which cannot be paired to their neighbours through these discrete gradients. Note that the required level of compliancy to achieve this construction has the virtue of not only producing the discrete set of critical segments, but also producing the triangulated walls and voids at no extra computational cost. The second step followed from the concept of the topological persistence (Edelsbrunner, Letscher & Zomorodian 2000, 2002), which assigns a density ratio to pairs of critical points which are found to be connected together by such discrete integral paths; these pairs are identified by the destruction/creation of critical points as one describes the level sets. If this ratio is below a given threshold, then the corresponding critical line/surface is found to be (topologically) insignificant, it is removed from the set, and the remaining critical mesh is simplified so as to recover some topological consistency. In other words, if the shot noise induces the occurrence of the discrete counterparts of, for example, spurious loops, disconnected blobs or tunnels which are found to be insignificant according to the aforementioned criterion, then they will be removed. This idea of the topological persistence was central in DisPerSE to produce a natural (topologically motivated) self-consistent criterion to infer the significance level of the identified structures. In particular, it warrants that the removal of pairs of critical points consistently extracts the corresponding topological feature (loop, tunnel, component, etc.). Note that those implemented theories are intrinsically discrete and readily apply to the measured raw data [modulo the consistent labelling of the elements of the Delaunay tessellation relative to the Delaunay

<sup>1</sup> Or any regular mesh as it happens.

<sup>2</sup> This well-known duality between the topology of the level sets and the characteristics of the critical points clearly has a discrete analogue through the creation/destruction of discrete cycles.

tessellation field estimator (DTFE) densities computed at the sampling points]. This warrants that all the well-known and extensively studied properties of the Morse theory are ensured by the construction *at the mesh level* and that the corresponding cosmological structures therefore correspond to well-defined geometrical objects with known geometrical and topological properties.

In some sense, DisPerSE filters out less-robust topological features, while keeping unchanged more robust ones, in analogy to what happens for the spectral linear low pass filtering. Instead of damping high frequencies, the persistence effectively ‘filters out’ from the critical sets (standing for the astrophysical filaments, walls, etc.) the topological features (loops, holes, etc.) which are less relevant, as they would most likely change or disappear when a different sampling or noising procedure is imposed on the underlying field. This ‘filtering’ process is non-linear and is formally equivalent to a scale-adaptive filtering where the scale would be decided by the local topology itself. As such, it is clearly more complicated to model than spectral filtering, but warrants the exact preservation of more robust topological features of the underlying field. This is clearly a requirement of any method which intends to, say, preserve the rough connectivity of the cosmic web. It is also a necessary requirement to preserve the underlying geometry.

Paper I presented a couple of examples of such a construction in 2D. Let us now illustrate the virtue of the method in the context of the 3D cosmic web. We start<sup>3</sup> by showing that the geometry of the cosmic web is accurately reproduced, while illustrating the quality of the cosmological structure identification, both in an  $N$ -body simulation (Section 2) and directly in an unprocessed version of the 7th Data Release (DR7) of the SDSS (SDSS DR7) galaxy catalogue (Section 3). In particular, we show how DisPerSE allows us to outline various configurations of the filamentary structure of galaxies and identify a previously missed X-ray ‘optically faint’ halo at the intersection of a set of SDSS filaments using the *Suzaku* satellite. We then discuss in Section 4 the problem of estimating the right value for the persistence level in cosmological simulations and illustrate how the measured topological properties of the LSS distributions are affected by varying this threshold. In particular, we show how this criterion compares with the simple friends-of-friends (FOF) algorithm when attempting to identify haloes in simulations. Section 5 wraps up and discusses prospects.

## 2 GEOMETRY OF THE LARGE-SCALE STRUCTURE: SIMULATION

We illustrated in Paper I how DisPerSE was able to correctly recover the geometry of the filamentary structure in the 2D case. We would like to show here how it performs with actual 3D cosmological data sets. However, demonstrating that a given algorithm is able to correctly identify the location of filaments is a difficult task, as it requires the previous knowledge of the location of those filaments. One possible solution therefore is to build an artificial distribution from a previously defined filamentary structure. This method was adopted in Aragón-Calvo et al. (2010b), where the authors use a Voronoi kinematic model (van de Weygaert 2002). We argue that using such a model to quantify the quality of the Morse–Smale complex identification is not as discriminatory as one would think (see Appendix A) and, for the lack of a simple and better way, we

<sup>3</sup> Note that our goal here is not to present an exhaustive review of the geometrical properties of the cosmic web, which is clearly out of the scope of this paper.

will therefore use here what is probably to date the most-efficient way to detect structures: the human eye and brain. Let us first start none the less with a little reminder on the notion of the persistence and significance threshold.

### 2.1 Persistence and significance threshold

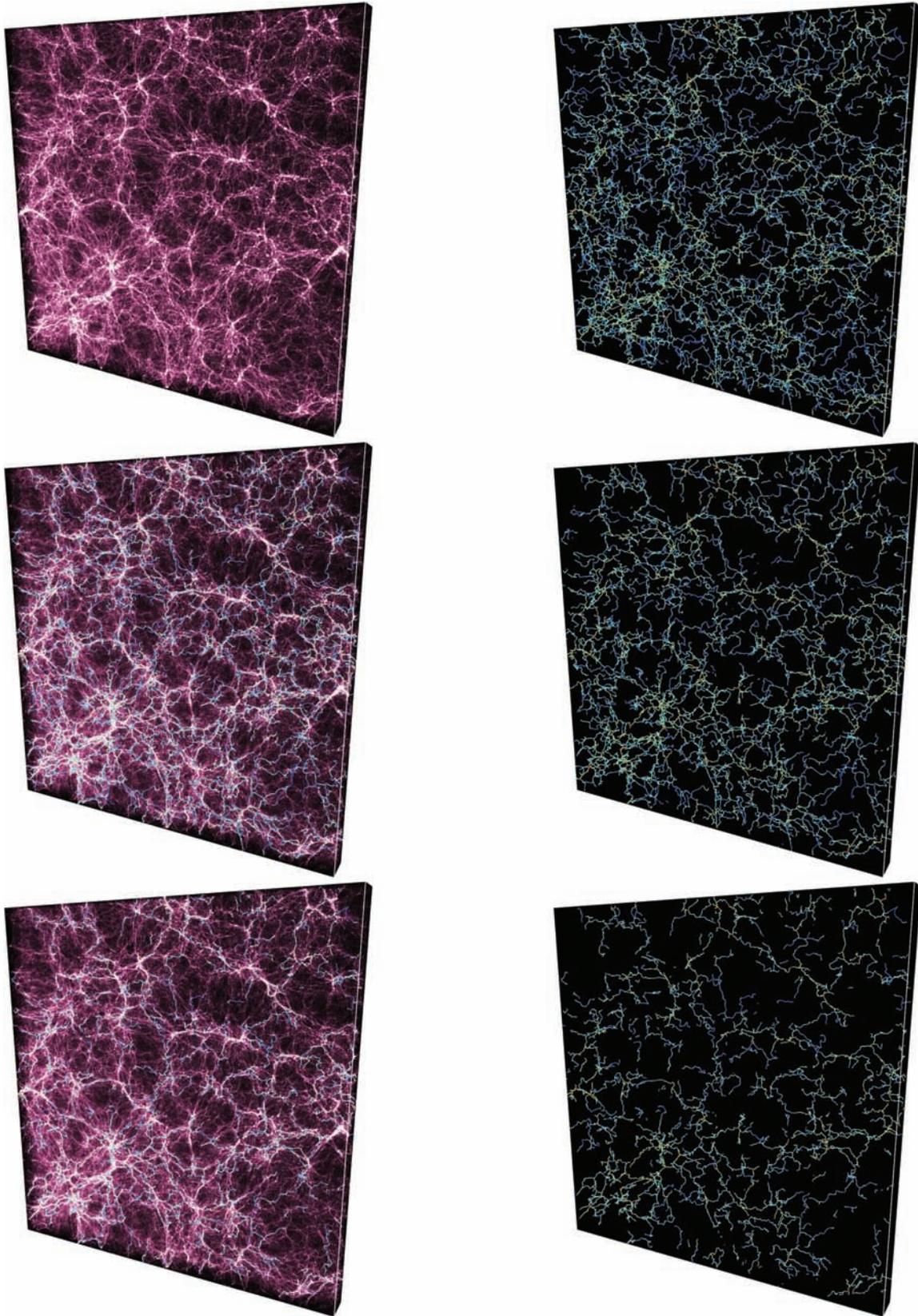
It was shown in Paper I how DisPerSE is able to measure topological properties of the cosmic web efficiently even in the presence of significant noise. This is achieved using a notion called the persistence. The concept is relatively simple: one counts the appearance of critical points in the excursion set of a function as the threshold defining the set is lowered. Some critical points entering the excursion set create topological components (for instance, a new isolated component is created whenever a maximum enters), while others destroy them (a saddle point may, for instance, link two previously isolated components, therefore destroying one of them). By pairing the two critical points that create and destroy a given feature into a persistence pair, one is therefore able to assess the ‘lifetime’ of the corresponding topological feature within the excursion. Often, the robustness of a topological feature – called its persistence – is measured as the difference in the density of the critical points in its persistence pair. Intuitively, it measures how much a function can be modified without possibly affecting a given topological feature. In the case of an underlying function with added white noise, for instance, the persistence would therefore allow one to easily measure the probability that a given feature would be a genuine characteristic of the underlying function or not at a given confidence level, expressed in terms of the variance of the noise,  $n\sigma$ .

In this paper, we will be dealing with discrete cosmological data sets: sets of particles sampling an underlying density function. DisPerSE deals with this sort of data by means of the Delaunay tessellation, which allows the reconstruction of the local topology of space (i.e. the properties of the neighbourhood of the sampling points) while giving a parameter-less and scale-free estimate of the density through the DTFE (Schaap & van de Weygaert 2000; van de Weygaert & Schaap 2009). The main source of the error on the estimate of the underlying density function is the Poisson noise that naturally results from the finite sampling. However, because of the scale-adaptive nature of the Delaunay tessellation, the way the Poisson noise affects our reconstruction (and in particular its topology) is not trivial. It was empirically measured in Paper I (see section 6.3) in terms of the probability that a persistence pair with the given persistence could be generated or destroyed by the Poisson noise only. This probability is better expressed in terms of the persistence ratio<sup>4</sup> (i.e. where the persistence of a pair is measured as the ratio of the density of the two critical points in the pair) and we will therefore refer in what follows to structures identified above a given persistence ratio, expressed as a number of  $\sigma$ ,  $n\sigma$ . By this, we mean that we are considering the structures that are characteristic of the underlying density field (i.e. not generated by noise) at a confidence level of  $n\sigma$  (levels of 1, 2, 3 and  $4\sigma$  corresponding to a probability of  $\sim 0.68$ ,  $\sim 0.95$ ,  $\sim 0.997$  and  $\sim 0.999\,937$ , respectively).

### 2.2 Visual inspection

The evolution of the geometry of the measured filaments with a significance threshold is illustrated in Fig. 1. The discrete Morse–

<sup>4</sup> Equivalently, one could of course consider the regular persistence – that is, density difference – of the logarithm of the density field.



**Figure 1.** The identified filamentary distribution of a  $512^3$ -particle and  $250 h^{-3} \text{ Mpc}^3$  large cosmological simulation at different significance levels in a  $250 \times 250 \times 20 h^{-1} \text{ Mpc}$  slice (above a persistence level of 3, 4 and  $5\sigma$ , from the top to bottom). The top left-hand panel shows the matter distribution, while the middle left-hand panel and bottom left-hand panel display the same distribution with filaments superimposed. The filamentary structure alone is displayed on the right-hand column. The computation was achieved on a  $128^3$ -particle subsample and the filaments are coloured according to the logarithm of the density. The density field was represented using the  $512^3$  particles of the  $N$ -body simulation.

Smale complex (DMC) represented in this figure was computed at the significance levels of 3, 4 and  $5\sigma$  (from the top to bottom) within a  $128^3$ -particle subsample of a  $512^3$ -particle,  $250 h^{-3} \text{Mpc}^3$   $\Lambda$  cold dark matter ( $\Lambda$ CDM) only  $N$ -body simulation. Note that the dark matter distribution within the  $20 h^{-1} \text{Mpc}$  slice is represented in the top left-hand corner to facilitate the visualization of its filamentary structure. Despite the projection effects that create visual artefacts (i.e. spurious filament-looking structures resulting from the projection of dark matter clumps at different depths) and the fact that filaments may enter or leave the slice, therefore seemingly appearing and disappearing for no apparent reasons, it seems fair to recognize that the agreement between the observed and measured filaments is excellent. These good performances are mainly the result of our use of the scale-adaptive Delaunay tessellation and the fact that our implementation does not require any pre-treatment of the density field, unlike usual grid-based methods which enforce a maximal resolution and resort to some kind of a density smoothing technique that affects the geometrical properties of the distribution. As a result, the resolution of the filaments is optimal with respect to the initial sampling, whatever be the selected significance level: the higher persistence and larger scale filamentary network is, by construction, a subset of its less-persistent and lower scale counterpart. Because a persistence-based topological simplification is used, increasing the persistence threshold actually results in less-significant filaments disappearing (when simplifying a 1-saddle point–2-saddle point persistence pair) or merging into each other (when simplifying a 1-saddle point–maximum persistence pair) to form larger scale more persistent ones, but conserving exactly the same resolution in any case. This can easily be observed by comparing the filamentary networks on the right-hand column of Fig. 1.

Another remarkable advantage of constructing a cosmological structure identification on the Morse theory is the extraordinary built-in coherence of the results, whatever be the type of the structure, as shown in Figs 2 and 3, for instance, the intricate pattern of a randomly selected void (i.e. an ascending 3-manifold) embedded within the filamentary network (i.e. ascending 1-manifolds) of the cosmic web in the same simulation as previously shown in Fig. 2. The location of the void within the slice is displayed in panel (b), each coloured square standing for a critical point (see legend). In the zoomed-in panels [panels (c) and (d)], the surface of the void has been shaded according to the logarithm of the density, showing how the DMC correctly traces the filamentary structure at the interface of the ascending 3-manifolds, as expected in the Morse theory.<sup>5</sup> Similarly, the neat relationship between a detected void and a wall structure on its surface (i.e. an ascending 2-manifold) in a  $100 h^{-1} \text{Mpc}$  large  $N$ -body simulation is displayed in Fig. 3.

Let us finally address a straightforward question: to what extent does DisPerSE manage to grasp the main features of the cosmic web with relatively sparse samples? Fig. 4 illustrates this query while comparing the filaments computed from two subsamples of varying resolution with a  $250 h^{-3} \text{Mpc}^3$  large cosmological simulation with  $512^3$  particles (namely  $64^3$ -particle and  $128^3$ -particle subsamples, respectively). From this figure, it seems that indeed the features which are identified in the sparser sample are real, since they are also found in the more densely sampled catalogue. There seems to be some encouraging level of convergence between the two sets of critical lines.

<sup>5</sup> The slight shift in the position between the surface of the void and the filament is due to the fact that we smoothed the filaments four times (see Paper I).

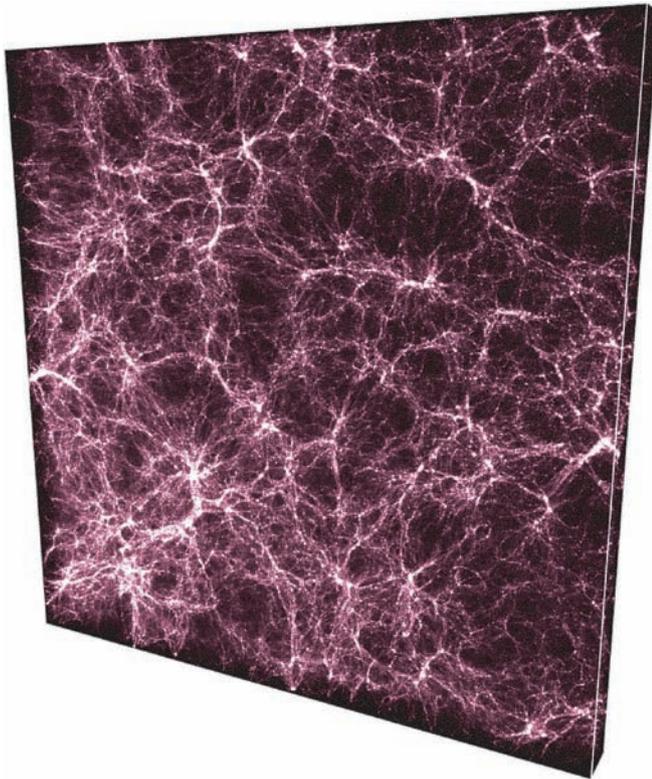
### 2.3 Persistent peak identification

From visual inspection, it therefore seems relatively clear that the technique developed in this paper is able to correctly decompose the cosmic web into simpler objects of astrophysical interest. However, the approach is based on one fundamental assumption, which is that the ascending and descending manifolds of the Morse theory, each associated to a specific type of critical point, are representative of the voids, filaments, walls and haloes. While the astrophysical nature of a filament or a wall is not defined very precisely, but is rather understood intuitively, this is not the case of a dark matter halo, for instance, which is supposed to be a gravitationally bound structure and the fact that the persistent maxima of the density field correctly identify the gravitationally bound structures is not established. We check this assumption by comparing the distribution of dark matter haloes identified by a simple FOF technique (see, for instance, Summers et al. 1995) in a  $100 h^{-3} \text{Mpc}^3$ ,  $512^3$ -particle  $\Lambda$ CDM simulation to the persistence diagram in the same simulation, as illustrated in Fig. 5.

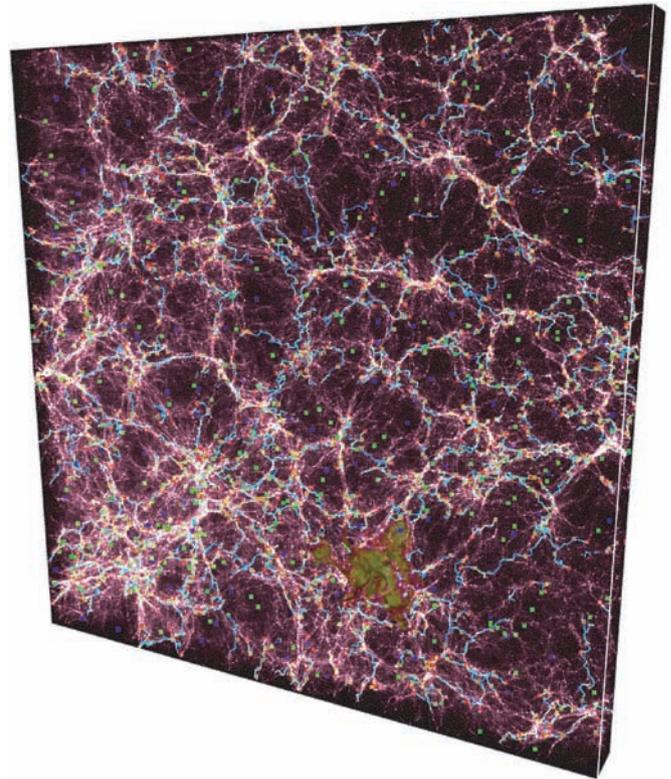
In this figure, the probability distribution function (PDF) of the persistence pairs<sup>6</sup> of type 2 (i.e. the maxima–1-saddle points pairs) measured in a  $128^3$ -particle subsample is displayed in the density–density plane, the horizontal axis corresponding to the density of the 1-saddle point and the vertical one to that of the maximum. The green line therefore represents the 0-persistence limit, while the oblique white dashed and dotted lines delimit the  $4\sigma$  and  $5\sigma$  thresholds, respectively. In order to compare this distribution to that of the astrophysical dark matter haloes, each of them is also represented as a disc with the coordinates that of the persistence pair of its most-dense particle (the densest particle within a halo is necessarily a local maximum). Each halo was identified using a standard linking-length parameter of one-fifth of the average interparticle distance and the red discs represent the haloes with mass  $M > 73.8 \times 10^{10} M_{\odot}$  (i.e. with more than 1280 particles in the initial simulation or 20 particles in a  $128^3$ -particle subsample), while the green ones represent the haloes with mass  $M > 590.4 \times 10^{10} M_{\odot}$  (i.e. with more than 10240 particles in the initial simulation or 1280 particles in a  $128^3$ -particle subsample). It is a very striking result how well the population of dark matter haloes is localized in the persistence diagram. While lighter ones (red discs) mostly correspond to maxima with the persistence ratio higher than  $4\sigma$  and overdensity  $\rho/\rho_0 > 4 \times 10^3$ , the heavier ones lie in the zone with the persistence ratio higher than  $5\sigma$  and overdensity  $\rho/\rho_0 > 3.2 \times 10^4$ .

These results mean that the persistence selection associated to a global overdensity threshold is naturally (i.e. without any specific calibration) a very good halo finder, which is quite encouraging, and validates our initial assumption on the relationship between the persistent topological features and the astrophysical components of the cosmic web. This is illustrated in Fig. 6 where each dark matter halo with mass  $M > 73.8 \times 10^{10} M_{\odot}$  (i.e. the red discs of Fig. 5) is coloured in blue. Once again, it is clear in the middle panel that all haloes along large filaments are correctly linked by the DMC. We also remark that the DMC and persistence pairs contain unexploited information of the topology as our algorithm explicitly identifies the  $k$ -cycles as sequences of critical points associated to persistence pairs (see Paper I). For instance, each persistence pair

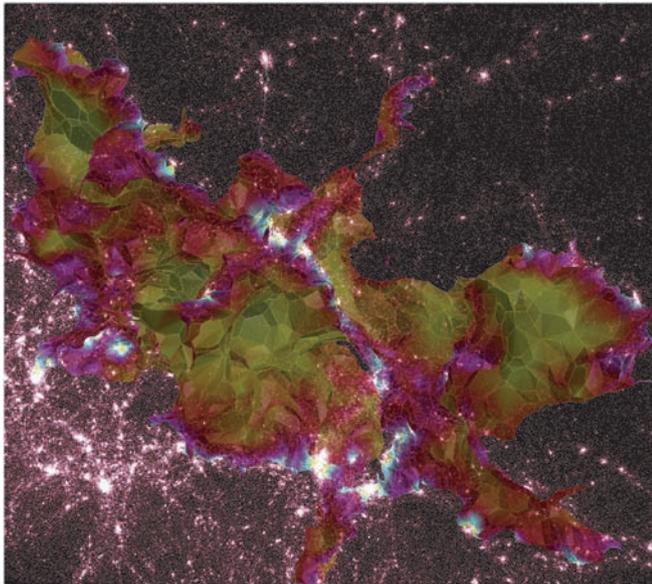
<sup>6</sup> As in Section 4, each pair is represented by a point with coordinates the density of each critical point within the pair, see that section for more explanations.



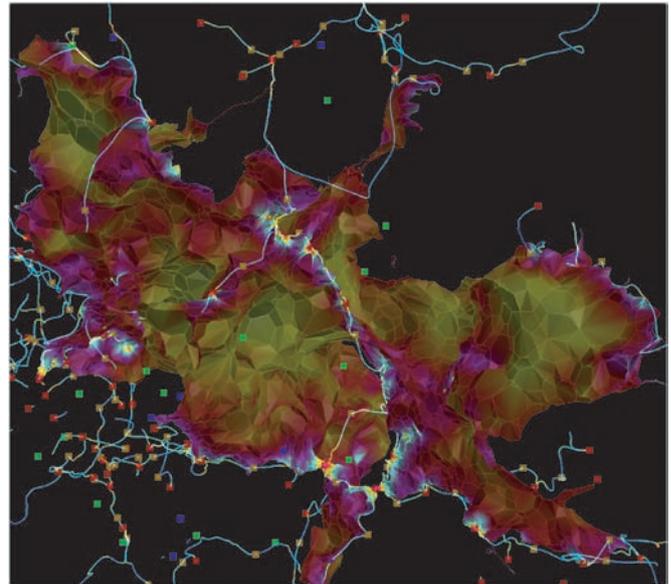
(a) Simulated dark matter distribution



(b) A void (bottom right) embedded in the filamentary structure

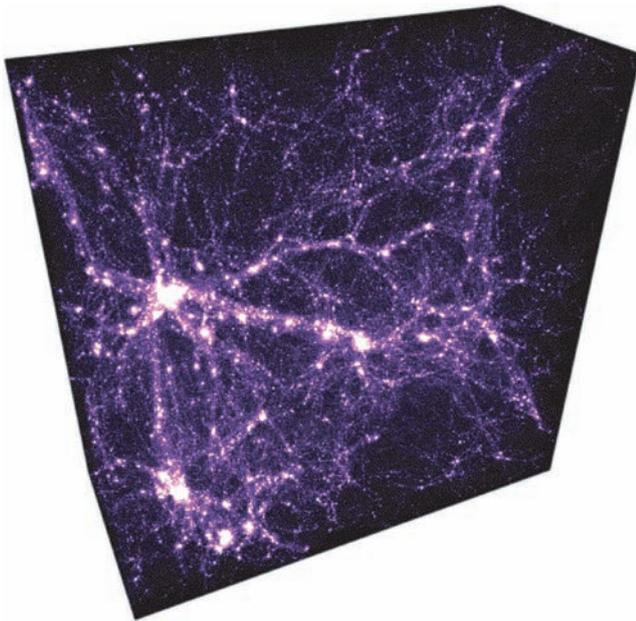
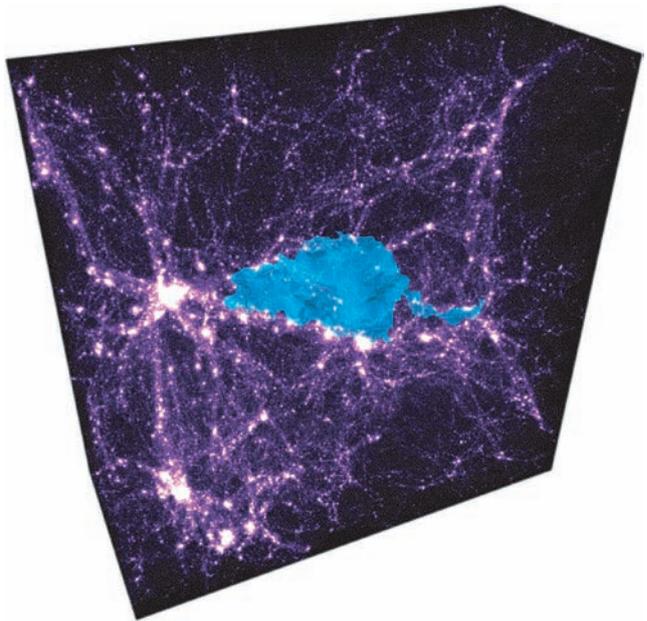
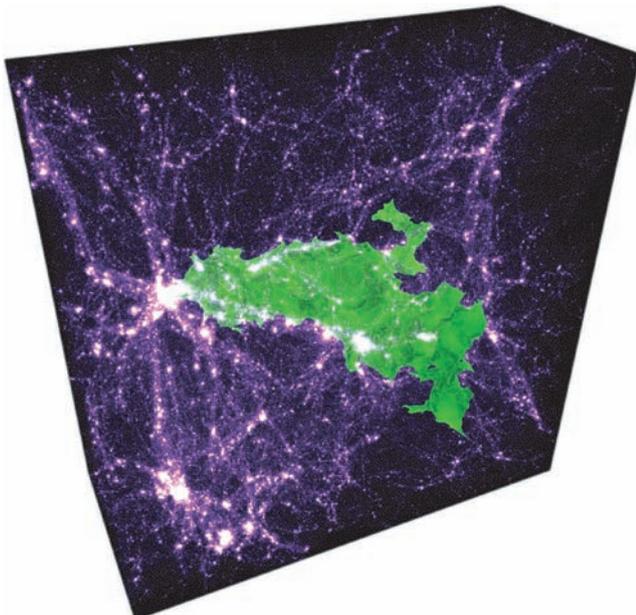
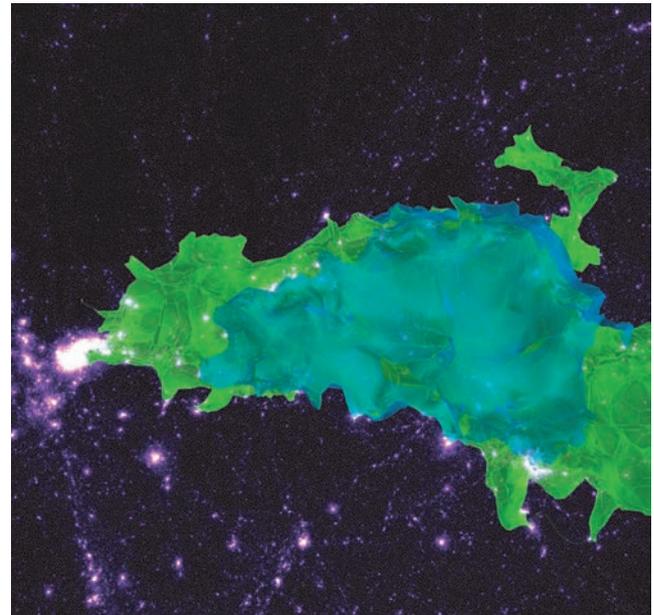


(c) Zoom on the void of panel 2(b).



(d) The relationship between the detected void, filaments and critical points

**Figure 2.** The arcs of the DMC (i.e. the filaments) and an ascending 3-manifold (i.e. a void) at a significance level of  $5\sigma$  in the same distribution as that of Fig. 1 (a  $250 \times 250 \times 20 h^{-1}$  Mpc slice of a  $512^3$ -particle and  $250 h^{-3}$  Mpc $^3$  large cosmological simulation). The density distribution is represented using all available particles within the simulation [panel (a)], while the DMC was computed using a  $128^3$ -particle subsample. The two lower panels [panels (c) and (d)] show a zoom-in on the upper panels at the location of the randomly selected void [see panel (b)]. On these figures, the maxima, 1-saddle points, 2-saddle points and minima are represented as the red, yellow, green and blue square, respectively, and the arcs as well as the manifold are shaded according to the log of the density. Note in panel (d) how the maxima, saddle points and path of the filaments correspond to the crests of the 2D density field measured on the surface of the void. This particularly emphasizes the coherence of the detection of objects of different nature.

(a) Dark matter distribution in a  $50 \times 50 \times 20 h^{-1}$  Mpc sub-box(b) An ascending 2-manifold (*i.e.* a wall)(c) An ascending 3-manifold (*i.e.* a void)

(d) Superposition of an ascending 3-manifold and an ascending 2-manifold on its surface.

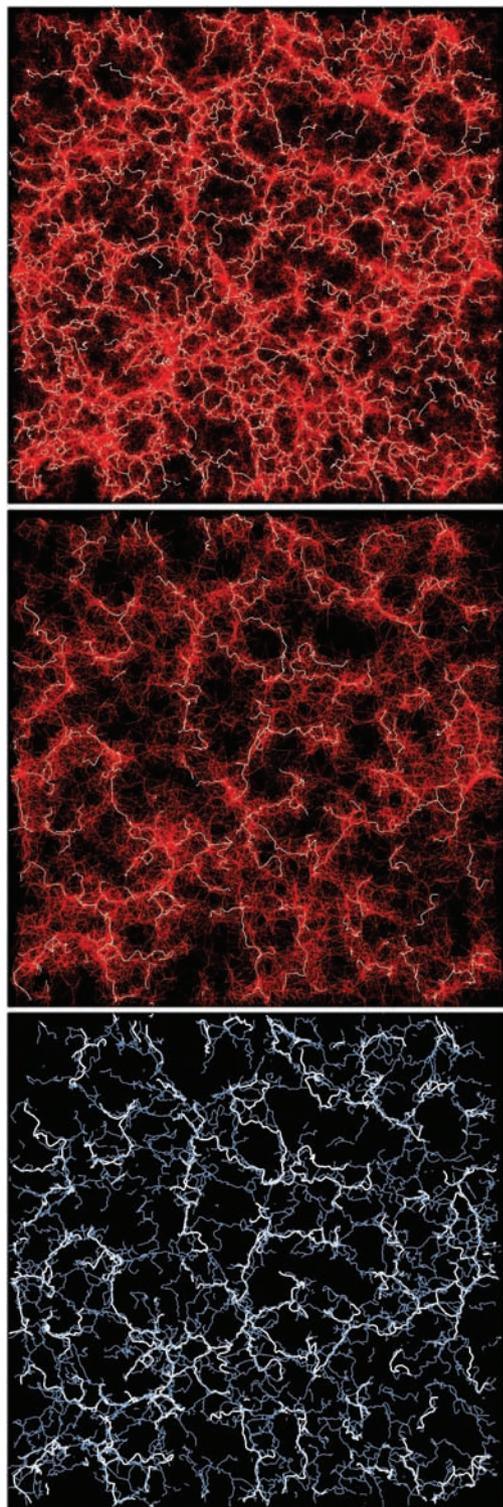
**Figure 3.** An ascending 2-manifold (*i.e.* blue 2D wall) and an ascending 3-manifold (*i.e.* green 3D void) identified in a  $512^3$ -particle,  $100 h^{-1}$  Mpc  $\Lambda$ CDM simulation. The manifolds were computed from a  $64^3$ -particle subsample.

associated to a halo corresponds to a 0-cycle that defines a principal filament, as shown on the bottom panel, where only the filaments corresponding to persistence pairs whose maximum is a dark matter halo are represented. Moreover, using the information contained in the persistence pairs, one basically obtains a hierarchical structure finder that is able to identify substructures not only within clusters, but also within filaments, walls and voids.

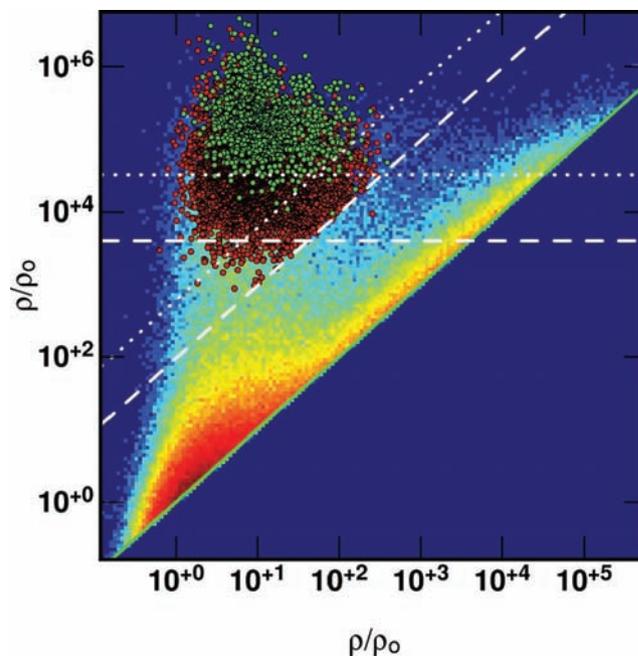
### 3 OUR UNIVERSE: THE SDSS CATALOGUE

Let us now illustrate a few prospective measurements of the filamentary structure of the actual galaxy distribution in the Universe.

The ultimate goal of such measurements is to allow a complete and precise characterization of the properties of the filamentary structure of the galaxy distribution by measuring their topological properties, such as the Betti number and Euler characteristics, and modelling the geometrical characteristics of the voids, walls and filaments (*i.e.* their total length, number, the number of filaments per galaxy cluster, etc.). Such a task is rather challenging, as it requires the construction of realistic mock observations from  $N$ -body simulations to assess the influence of observational biases and distortions; it also requires a lot of care in the handling of the observational data themselves (for instance, by taking into account the complex survey geometry, among other difficulties). In this paper, we focus



**Figure 4.** The filamentary distribution above a persistence level of  $4\sigma$  in a  $250 \times 250 \times 20 h^{-1}$  Mpc slice of a  $512^3$ -particle,  $250 h^{-3}$  Mpc<sup>3</sup> large cosmological simulation. The red segments in the top and middle panels correspond to the segments of the Delaunay tessellation of a  $128^3$ - and  $64^3$ -particle subsample, on which the corresponding filaments have been computed. In the bottom panel, the thick white filaments correspond to the  $64^3$ -particle subsample, while the blue thin filaments were computed on the  $128^3$ -particle subsample. This figure clearly demonstrates that DisPerSE is able to grasp the main features of the cosmic web with a relatively sparse sample.



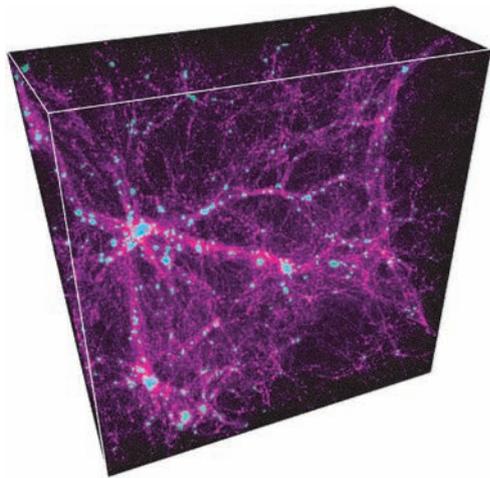
**Figure 5.** Distribution of the persistence pairs of the highest-density particles within each dark matter halo of mass  $M > 74 \times 10^{10} M_{\odot}$  (red) and  $M > 590 \times 10^{10} M_{\odot}$  (green) in a  $128^3$ -particle subsample of a  $100 h^{-1}$  Mpc large  $\Lambda$ CDM dark matter simulation. The density of the lowest critical point in a persistence pair is represented by its  $X$ -coordinate, while the density of its highest critical point is shown by its  $Y$ -coordinate [i.e. a point with coordinate  $(\rho_x, \rho_y)$  represents a persistence pair whose critical points have the corresponding densities]. The persistence diagram of maxima–1-saddle points pairs with the persistence larger than  $3\sigma$  is shown in the background. The horizontal dashed and dotted lines correspond to overdensity levels of  $4 \times 10^3$  and  $3.2 \times 10^4$ , respectively, and the oblique lines correspond to the persistence levels of  $\sim 4$  and  $\sim 5\sigma$ , respectively.

on convincing the reader that the method we introduced in Paper I is particularly suited to such a task by showing how easily and efficiently it can be applied to a real galaxy catalogue. We postpone the full investigation to a future paper.

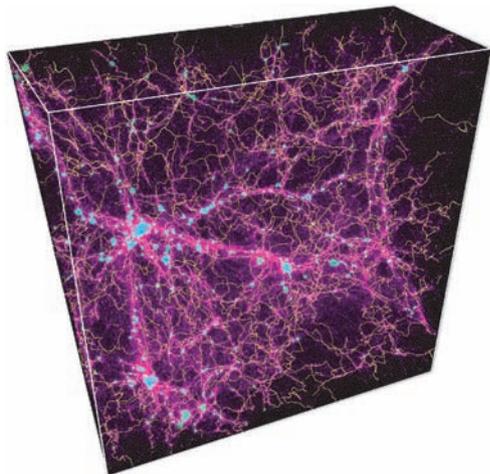
### 3.1 The cosmic web in the SDSS

For that purpose, we use SDSS DR7 data (Abazajian et al. 2009) and in particular the LSS subsample called the *dr72bright0* sample of the New York University Value Added Catalogue (Blanton et al. 2005), which is made of a spectroscopic sample of galaxies with  $u$ -,  $g$ -,  $r$ -,  $i$ -,  $z$ -band ( $K$ -corrected) absolute magnitudes, an  $r$ -band apparent magnitude  $m_r$ , redshifts and information on the mask of the survey. In that sample, the spectroscopic galaxies are originally selected under the conditions that  $10.0 \leq m_r \leq 17.6$  and  $0.001 \leq z \leq 0.5$ , but we further cut the sample for the purpose of our tests, restraining it to the galaxies with  $z \leq 0.26$  and  $100^\circ \leq \text{RA} \leq 280^\circ$ , which removes the three thin stripes in the Southern Galactic hemisphere. The resulting angular distribution, containing 515 458 galaxies out of 567 759 in the original sample, is displayed in Fig. 7.

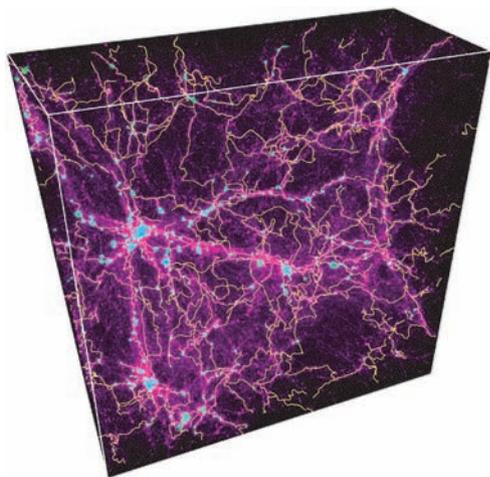
In order to compute the DMC of the observed galaxy distribution, we will use the mirror-type boundary conditions as introduced in Paper I. This type of boundary conditions normally apply to distributions enclosed within parallelepiped boxes, which is not the case here. In the simple case of a box-like geometry, the particles within a given distance of the faces are mirrored and any particle



(a) Dark matter distribution in a  $50 \times 50 \times 20 h^{-1} \text{Mpc}$  sub-box and haloes with mass  $M > 73.8 \cdot 10^{10} M_{\odot}$

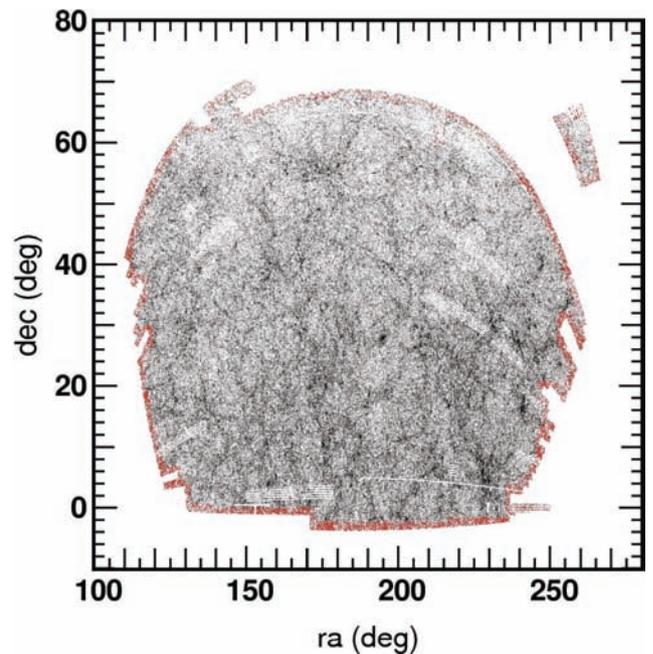


(b) The filaments at  $4\text{-}\sigma$  on a  $128^3$  sub-sample



(c) The main filaments of the dark matter haloes

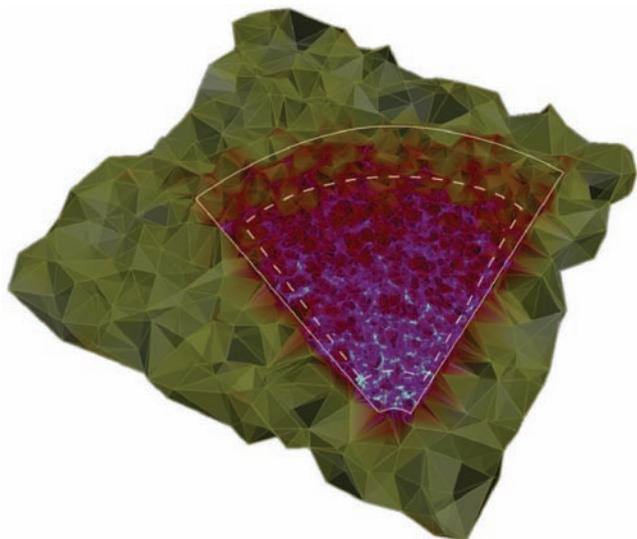
**Figure 6.** Distribution of the main filaments of FOF haloes with mass  $M > 73.8 \cdot 10^{10} M_{\odot}$  in a  $20 h^{-1} \text{Mpc}$  thick slice of a  $512^3$ -particle,  $100 h^{-3} \text{Mpc}^3$   $\Lambda\text{CDM}$  dark matter simulation. The filaments were computed from a  $128^3$ -particle subsample. Note that many filaments are linked to haloes outside the slice, giving the false impression to end for no reason.



**Figure 7.** Angular distribution of the 515 458 galaxies corresponding to a subsample of the SDSS DR7 galaxy catalogue that we use in our tests (see the main text for the selection criterion). The 66 608 red galaxies are those detected as being on the boundary of the distribution using the method described in the main text. Note that some regions were not fully scanned and exhibit a series of thin empty parallel stripes, but we simply ignore that fact when computing the boundaries.

outside the initial box or whose DTFE density may be affected by the content of the exterior of the box is tagged as a boundary particle. As the geometry of the SDSS catalogue is complex, we simply enclose it within a slightly larger box, fill the empty regions with a low-density random distribution of particles and then mirror the boundaries. The mirrored particles and the random ones are tagged and we then identify the boundaries of the galaxy distribution and tag as well those galaxies whose DTFE density may depend on the distribution outside the observational region. Although the catalogue does contain precise information about the mask of the survey, we prefer to use a simple though automatic method to identify the boundaries of the galaxy distribution. This method simply samples the angular galaxy distribution in the RA–Dec. plane over a regular grid of  $1^\circ \times 1^\circ$  pixels and identifies the galaxies on the boundary of the catalogue as those that belong to a pixel with at least one completely empty neighbour. Note that such a method presents the advantage of being generic, as it does not presume any previous knowledge of the mask, and could therefore be applied directly to other galaxy catalogues. The resulting boundary galaxies are represented in red in Fig. 7. Finally, we also tag those galaxies with redshifts  $z \leq 0.02$  and  $z \geq 0.2$  as the boundary and proceed with the computation of the DMC, as in the regular mirror-type boundary condition case. A slice of the Delaunay tessellation of the final distribution is displayed in Fig. 8.

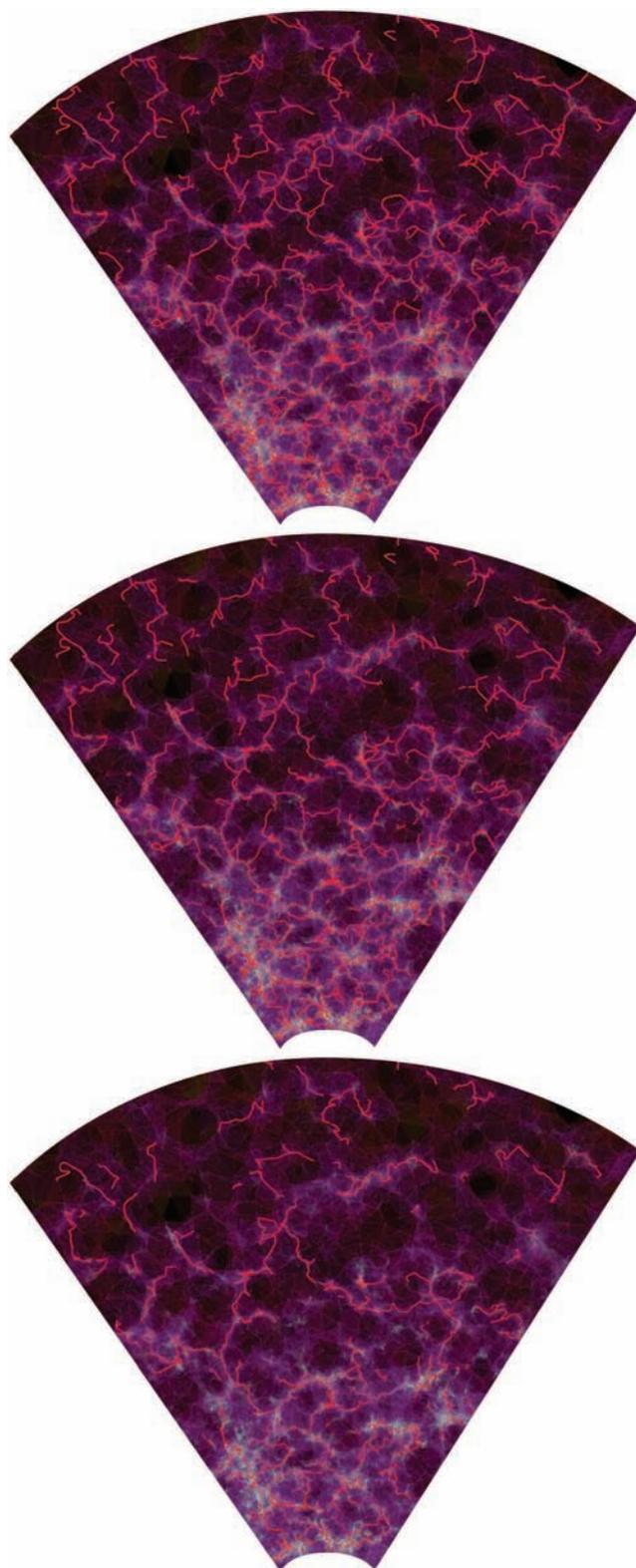
The resulting DMC covers the 440 950 galaxies in black in Fig. 7 by obeying the additional condition  $0.02 \leq z \leq 0.2$  (or equivalently  $85 \leq d \leq 860 h^{-1} \text{Mpc}$ ) and it is displayed in Figs 9–11. Fig. 9 illustrates the influence of the significance level on the measured filamentary network. In this figure, the filaments (i.e. the ascending 1-manifolds or arcs) within a  $\sim 40 h^{-1} \text{Mpc}$  slice of the Delaunay tessellations are shown at the significance levels of 3, 4 and  $5\sigma$



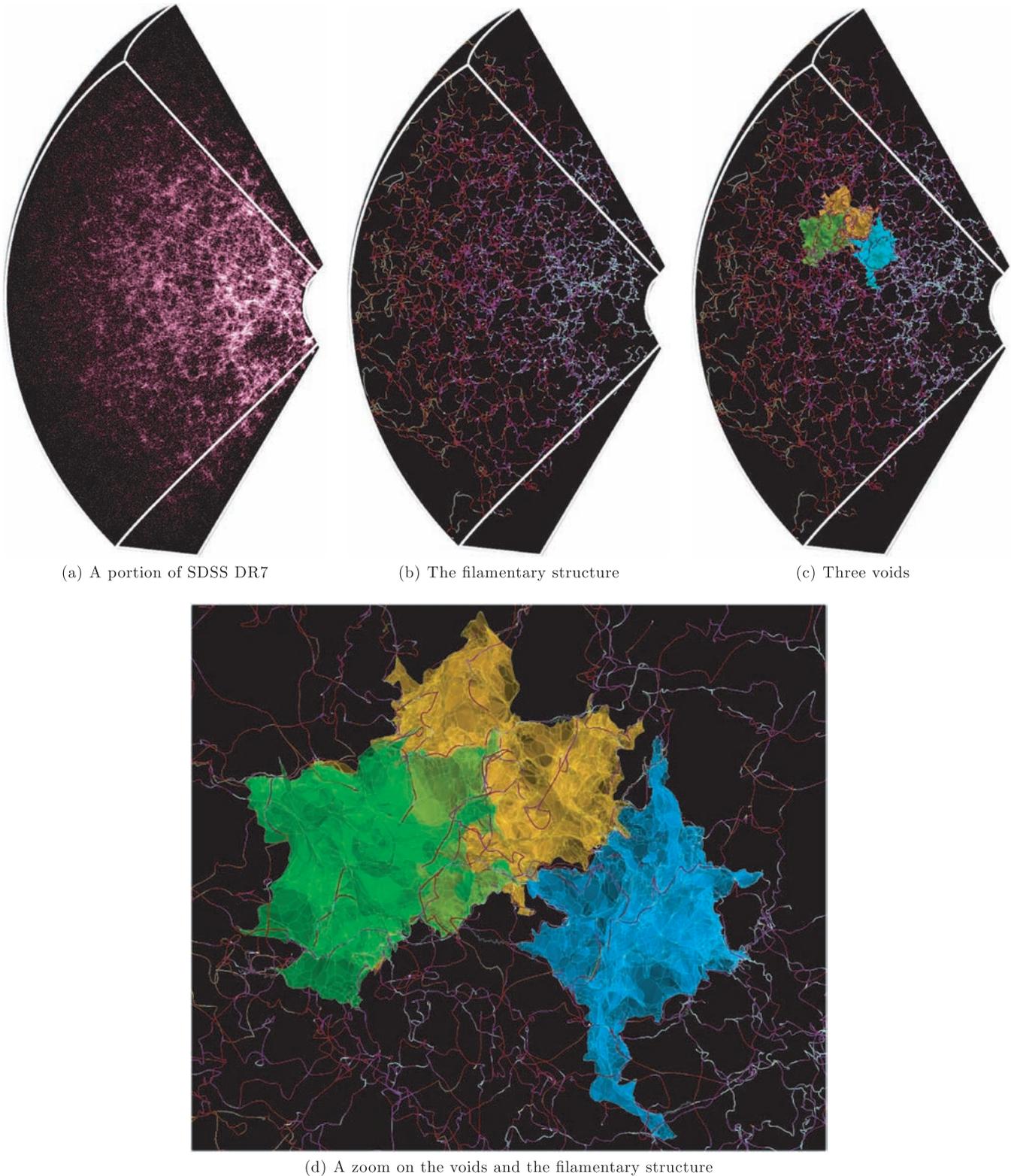
**Figure 8.** A slice within the Delaunay tessellation of the distribution used to compute the DMC of the SDSS. The plain white contour delimits the SDSS distribution (inside) and the randomly added low-density particles that fill the void regions of the bounding box (outside). Any galaxy outside the white dashed contours is considered as being on the boundary.

(from the top to bottom); it is quite striking how well more- or less-significant filaments are accurately identified depending on the value of the persistence ratio threshold. Note how already at a level of  $3\sigma$  the influence of the sampling noise has disappeared and increasing this threshold results in the selection of apparently denser, bigger and longer filaments. As the distant faint galaxies and the nearby bright ones cannot be observed easily, the selection function strongly depends on the distance and so does the sampling. It reflects in the shade of the Delaunay tessellation, which depends on the logarithm of the density. From a theoretical point of view, the fact that the absolute value of the density is multiplied by the selection function should not affect the detection of the filaments as long as the value of the selection function does not vary much from the typical scale of a filament (or in other words, as long as the topology of the distribution remains unchanged). The measured persistence ratio of persistence pairs may be slightly affected though, when the two critical points in the pair are located at different distances, but this does not seem to have much importance in the present case. A more significant effect results from the scale-adaptive nature of the DTFE. Because the quality of the sampling decreases with the distance, comparatively larger scale filaments are identified as the distance increases and to be able to identify comparable filaments independently of the distance from the observer, one would therefore probably have to resort to volume-limited samples.

The filamentary structure at the  $5\sigma$  significance level is also shown over larger scales in Fig. 10 and within a  $60 h^{-1}$  Mpc slice where each galaxy is represented by a point in Fig. 11. Three voids (i.e. ascending 3-manifolds) have been randomly selected within the distribution of Fig. 10 and are displayed in the bottom panel [panel (d)], showing the intricate relationship between the voids and the filamentary structure that crawls at their surface. As previously observed in simulations, it can be seen in the middle right-hand panel of Fig. 11 that those 3D filaments also trace the 2D filamentary structure at the surface of the voids as expected from the Morse theory. Note that it is only because they have been smoothed over four



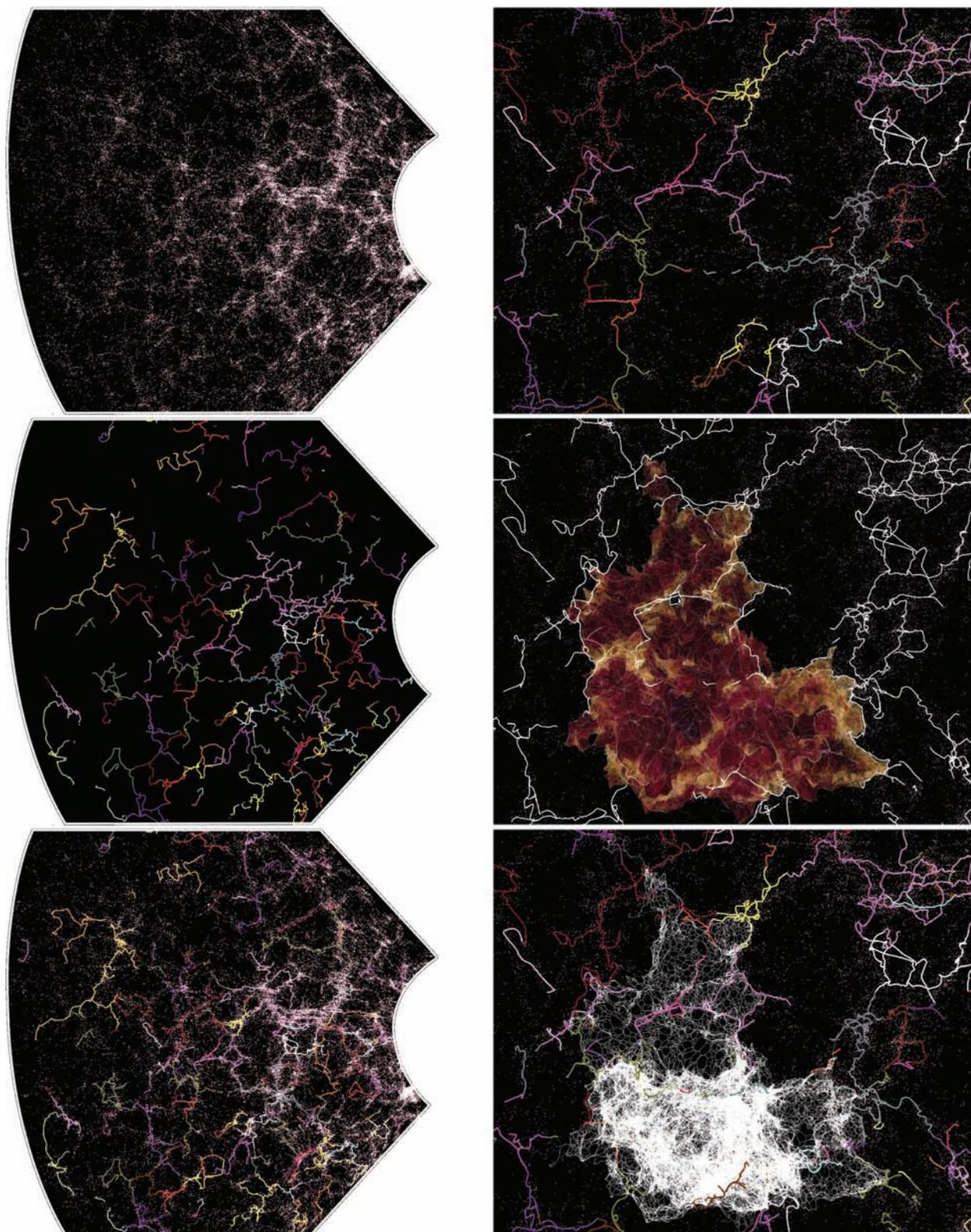
**Figure 9.** From the top to bottom, the filamentary structure in a  $\sim 40 h^{-1}$  Mpc thick slice of the SDSS DR7 galaxy catalogue at a significance level of 3, 4 and  $5\sigma$ , respectively. The distribution is represented by the non-bounding subset (see main text) of the Delaunay tessellation used to compute the DMC, shaded according to the logarithm of the density. The depth of a filament can be judged by how dimmed its shade is. Note that filaments that seem to stop for no apparent reason actually enter or leave the slice.



**Figure 10.** The detected filamentary structure at a significance level of  $5\sigma$  and three voids within a portion of the SDSS DR7. Note that only the upper half of the distribution shown in Fig. 7 is displayed here for clarity reasons. The colour of the filaments corresponds to the logarithm of the density field.

segments to look more appealing and to avoid rendering problems that the filaments do not lie precisely on the surface of the voids. It is in fact a built-in feature of the DMC and in particular of our implementation that all the different types of identified cosmological structures do form a coherent picture, whatever be the properties of

the initial discrete sample. This allows for interesting features, such as making possible the count of the number of filaments that belong to a common maximum by intersecting the ascending 1-manifolds with the descending 3-manifolds. This is shown in Fig. 11 where the colour of the filamentary structure corresponds to the index of the



**Figure 11.** The filamentary structure (left-hand panel) and a void (right-hand panel) detected at a significance level of  $5\sigma$  in the SDSS DR7. In order to emphasize the filamentary structure, only a  $\sim 60 h^{-1}$  Mpc thick flat slice of the distribution is displayed in each panel. The void surface is shaded according to the log of the density field (middle right-hand panel), while the colour of each arc of the DMC tracing the filamentary structure depends on the index of the maximum to which it is connected. Note that the foremost part of the voids in the middle and bottom right-hand panels protrudes from the slice, while the filaments are trimmed to its surface. Given its shape, this void is in fact a good example of why we should identify filaments via a DMC rather than using the Watershed technique, as it displays two strong ‘thin wings’ which would lead to the incorrect detection of spurious sets of boundaries.

maximum it belongs to and individual filaments could be identified the same way, as the two arcs of the DMC originating from a given saddle point.

### 3.2 An ‘optically faint’ cluster at a filamentary junction

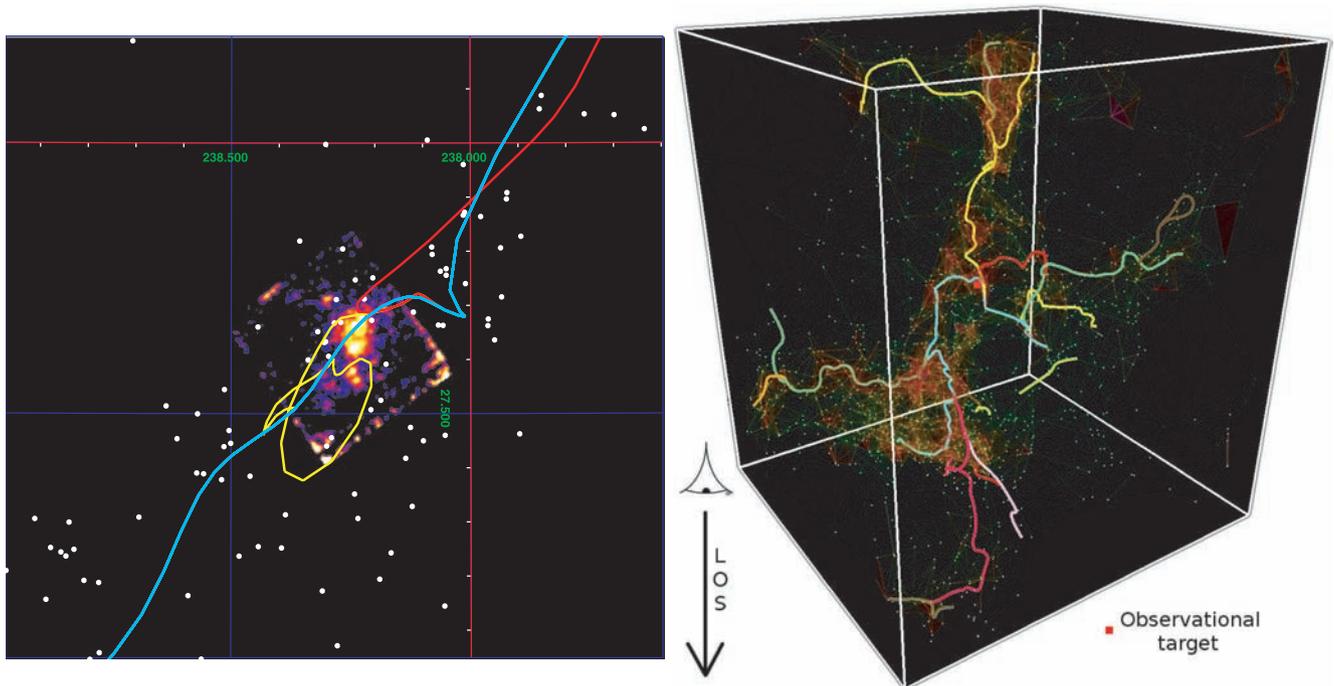
Because some dark matter haloes are sparsely populated and also as a result of selection effects, classical methods, such as FOF, are unable to detect them in the observed galaxy distribution. Such ‘optically faint’ groups and clusters may nevertheless present a strong astrophysical interest: as they are by nature different from the ‘regular’ haloes, one could, for instance, expect that they have different formation history that needs to be understood. As they are faint though, their properties are poorly assessed, but massive dark matter haloes such as galaxy clusters or galaxy groups are believed to form at intersections of two or several filaments, which can be identified in the SDSS using DisPerSE. We demonstrate that this is possible by enlightening the relationship between an X-ray halo and its surrounding filamentary network as identified in the SDSS catalogue (see Fig. 10b).

Because of the particular configuration of the filaments in the region, we submitted an observation proposal to the X-ray satellite *Suzaku* (Mitsuda et al. 2007), which was accepted. We show here the results of this observation, but the analysis is presented in Kawahara et al. (2011). The observational target was selected for being located at the confluence of galaxy filaments and because one of those filaments is both straight and aligned with the line of sight as shown in Fig. 12 (see the yellow filament in the right-hand panel). While no

X-ray signal could be found within the *ROSAT* All Sky Survey, X-ray signals emitted by the diffuse thermal gas were clearly observed by the high-sensitivity detectors of *Suzaku*, unveiling the presence of a dark matter halo as shown by the X-ray image reproduced in the middle part of the left-hand panel of Fig. 12. It is remarkable that there are no corresponding candidates in the 78 800 group catalogue identified by Tago et al. (2010) using a modified FOF algorithm. In fact, because the optically observable member galaxies are not strongly clustered and their number is limited ( $N \sim 10$ ), regular methods have high chances to miss them. It is also very difficult to locate and identify particular filamentary configurations by eye directly from the galaxy distribution using projections or even a real-time 3D visualization. Using DisPerSE, we showed that it is possible to easily identify such targets, which demonstrates the complementarity of our approach with respect to one based on a traditional halo finder.

## 4 SIGNIFICANCE OF THE TOPOLOGY OF THE LARGE-SCALE STRUCTURE

As noted in Paper I, it is not an option to use the raw DMC as a tool to assess the properties of the cosmic web. Hence, we showed there how to simulate a topological simplification of the DTFE density field so that the critical simplexes that were most probably accidentally generated by the Poisson noise could practically be removed from the DMC. This simplification is based on the persistence ratio of critical point pairs (i.e. persistence pairs) and one must therefore decide a significance level  $s = n\sigma$  such that all persistence

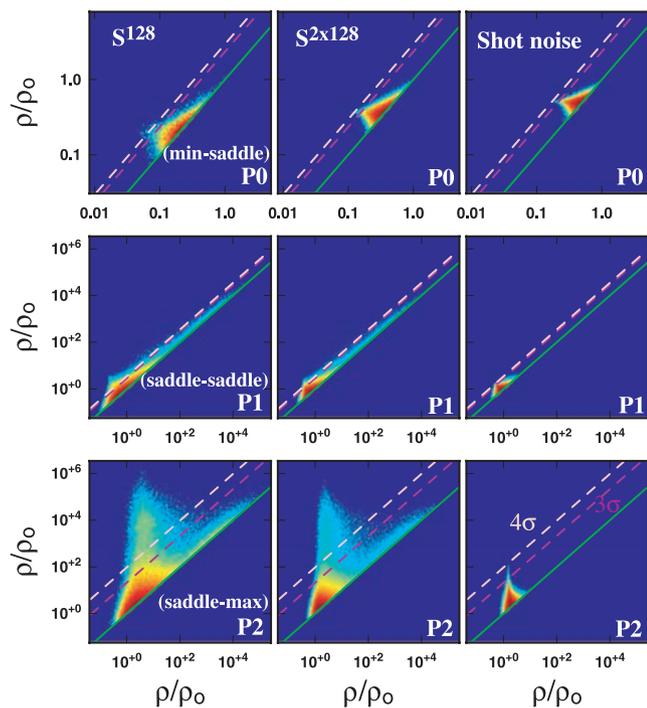


**Figure 12.** Left-hand panel: an X-ray halo observed around an elliptical galaxy in the centre of a group at redshift  $z = 0.083$  and located at the confluence of several filaments. The colour map indicates the X-ray-combined image of CCD chips (XIS 0, 1 and 3), while the white dots stand for the SDSS spectroscopic identified galaxies within  $0.080 < z < 0.086$ . The filamentary structure in the surrounding region is shown by the coloured solid curves, extracted from the filament catalogue shown in Fig. 10. Note that the colours (cyan, red and yellow) correspond to that of the filaments represented on the 3D view on the right-hand panel. Right-hand panel: a 3D view of the configuration of the filaments around the observed region. The vertical axis corresponds to the line of sight (the observer being upwards) and the box roughly encompasses the galaxies in the SDSS catalogue with coordinates  $233^\circ < \text{Dec.} < 243^\circ$ ,  $22^\circ < \text{RA} < 32^\circ$  and  $0.075 < z < 0.092$ . The Delaunay tessellation of the galaxies, shaded according to the local density, is displayed to help visualize the filamentary structure. The observational target is identified by a red square and is located at the intersection of the red cyan and yellow filaments, the last two being aligned with the line of sight to a very good approximation. A movie is available for download at <http://www.iap.fr/users/sousbie/>

pairs with lower significance (or equivalently a higher probability to be generated by the Poisson noise) can be removed. We showed in Paper I that, at least in the 2D case, such a method allows for what seems to be a very efficient and natural simplification of the DMC. We did not discuss, however, how to decide the value of this particular threshold. This is particularly important though, and especially in the context of the cosmic web, as our ultimate goal is to assess physical properties of astrophysical objects identified as features of the DMC (i.e. the haloes, filaments, walls and voids of the matter distribution on cosmological scales in the Universe). Imagine, for instance, one is interested in statistically measuring the average number of filaments that branch on dark matter haloes. If the threshold is too low, the measure will be equivalent to that in a Gaussian random field because of the Poisson noise (see the lower left-hand panel of fig. 13 of Paper I) and if it is too high, then the risk is to systematically ignore weaker filaments (see the middle right-hand panel of fig. 13 of Paper I).

#### 4.1 Persistence diagrams

Fig. 13 shows the PDF of persistence diagrams (see Edelsbrunner et al. 2000; Cohen-Steiner, Edelsbrunner & Harer 2007) computed



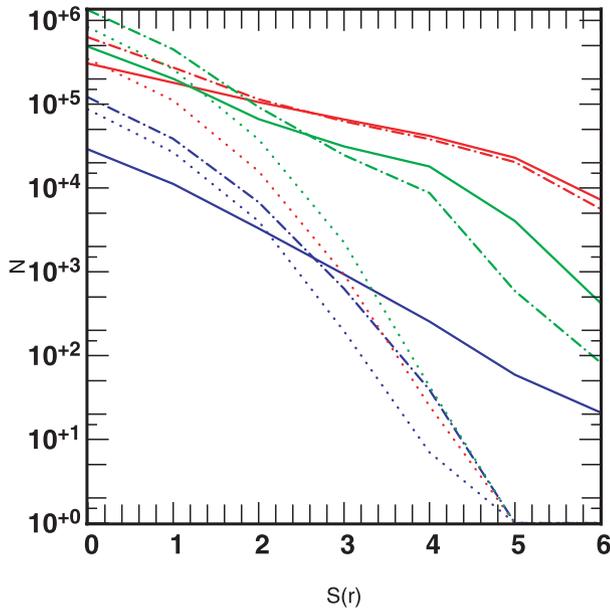
**Figure 13.** Persistence diagrams (i.e. the PDF of persistence pairs) in a cosmological simulation and for the Gaussian random noise. Each pair  $P_i = [p_i, q_{i+1}]$  of critical points of order  $i$  and  $i + 1$  is considered as a point with the coordinates  $[\rho(p_i), \rho(q_{i+1})]/\rho_0$ . The density of the lowest point in the pair (or equivalently of that with the lowest critical index) corresponds to the X-coordinate and that of the highest (or equivalently of that with the highest critical index) corresponds to the Y-coordinate. The PDF was computed from a  $250 h^{-1}$  Mpc large  $\Lambda$ CDM simulation downsampled to  $128^3$  particles,  $S^{128}$  (left-hand column), the same distribution with  $128^3$  additional randomly located particles,  $S_N^{2 \times 128}$  (middle column) and a random distribution of particles within the same volume,  $S_R^{128}$  (right-hand column). From the top to bottom, each line corresponds to a different type of a pair:  $P_0$  (minima–2-saddle points),  $P_1$  (2-saddle points–1-saddle points) and  $P_2$  (1-saddle points–maxima), respectively. The green, purple-dashed and pink-dashed lines correspond to the 0, 3 and  $4\sigma$  persistence levels, respectively.

from the Delaunay tessellation of a  $250 h^{-1}$  Mpc large,  $512^3$ -particle  $\Lambda$ CDM simulation subsampled to  $128^3$  particles (left-hand column; hereinafter  $S^{128}$ ), the same distribution with an identical number of particles added at random locations (middle column; hereinafter  $S_N^{2 \times 128}$ ) and a completely random distribution of particles within the same volume (right-hand column, hereinafter  $S_R^{128}$ ). Plotting a persistence diagram of a density distribution  $\rho$  basically consists in representing each persistence pair  $P_i = [p_i, q_{i+1}]$ , where  $p_i$  and  $q_{i+1}$  are the critical points of order  $i$  and  $i + 1$ , respectively, as a point with coordinates  $[\rho_{\downarrow}, \rho_{\uparrow}] = [\rho(p_i), \rho(q_{i+1})]/\rho_0$ , where  $\rho_0$  designates the average density in the distribution.<sup>7</sup> In Fig. 13, the pairs of type  $P_0$ ,  $P_1$  and  $P_2$  are represented in the top, middle and bottom rows, respectively. In those diagrams, the pairs with the null persistence lie on the green line of equation  $\rho_{\uparrow} = \rho_{\downarrow}$  and the farther away from this line a point is, the higher the persistence of its corresponding persistence pair. The purple and pink dashed lines stand for the 3 and  $4\sigma$  persistence, respectively. Recall that persistence pairs are pairs of critical simplexes that correspond to the act of the creation and destruction of a topological feature (i.e. component, loop, shell, etc.) in the filtration of the Delaunay tessellation. This basically means that each point in the diagram for  $P_0$ -,  $P_1$ - and  $P_2$ -type pairs (i.e. top, middle and bottom rows) stands for a filament, wall or void that could be considered a physical object below the persistence threshold of its representing pair but would not exist or be a part of a larger, more persistent object at higher persistence threshold. Filaments are represented by  $P_0$ - and  $P_1$ -type pairs (one needs a filament to close a loop in the filtration), walls by  $P_1$ -type pairs only and voids by  $P_1$ - and  $P_2$ -type pairs. A highly persistent  $P_0$  pair, for instance, therefore represents a very significant filament composed of many less-significant filaments, each represented by a lower persistence  $P_0$  or  $P_1$  pair.

As expected, most persistence pairs in the random distribution  $S_R^{128}$  have a persistence ratio below  $3\sigma$  (right-hand column). Fortunately, the PDF of the persistence pairs in  $S^{128}$  is sufficiently different from that in  $S_R^{128}$  so that a reasonable fraction of them lie above the  $3\sigma$  and even the  $4\sigma$  threshold (left-hand column). By canceling all those pairs that lie below the 3 or  $4\sigma$  line, it should therefore seem reasonable to assume that only those topological properties that were imprinted by the physical processes at work in the simulation would be conserved. A good measure of the actual influence of the Poisson noise on the distribution of the persistence pairs in the underlying distribution can be gained from the examination of the middle column. The distribution  $S_N^{2 \times 128}$  was created by adding a large number of randomly located particles to  $S^{128}$ , resulting also in the creation of a very large number of spurious critical points. One can see in the middle column that as a result the persistence diagram tends to concentrate at a lower persistence ratio (i.e. closer to the green line). This means that as expected those spurious critical points mainly create low-persistence-ratio pairs which can therefore be easily removed.

This observation is supported by Fig. 14, where the actual number of persistence pairs in the three distributions is displayed as a function of the cutting threshold, whereas the number of critical pairs of all sorts and with the significance higher than  $0\sigma$  is higher in  $S_N^{2 \times 128}$  (dot-dashed curves) than in  $S^{128}$  (plain curves); this number decreases comparatively faster with the increase in the persistence selection threshold. For low-persistence thresholds (i.e. up to  $\sim 2\sigma$ ), the number of persistence pairs in  $S_N^{2 \times 128}$  actually decreases as fast

<sup>7</sup> In the following, the term density will generally refer to the normalized density  $\rho/\rho_0$  so that different distributions can be fairly compared.



**Figure 14.** Number of persistence pairs of type  $k$  as a function of the significance threshold  $S_k(r)$  (in units of  $\sigma$ ) in a  $250 h^{-1}$  Mpc large  $\Lambda$ CDM simulation downsampled to  $128^3$  particles,  $S^{128}$  (filled curves), the same distribution with  $128^3$  additional randomly located particles,  $S_N^{2 \times 128}$  (dot-dashed curve) and a random distribution of particles within the same volume,  $S_R^{128}$  (dotted curves). The blue, green and red colours correspond to persistence pairs of type 0, 1 and 2, respectively (see Fig. 13 for the corresponding persistence diagrams).

as that in the random distribution  $S_R^{128}$  (dotted curves). In the case of pairs of type  $P_1$  and  $P_2$  (2-saddle point–1-saddle point pairs, green curves, and 1-saddle point–maxima pairs, red curves, respectively), this tendency actually changes between  $2-3\sigma$  and the cancellation rates in  $S_N^{2 \times 128}$  and  $S^{128}$  become relatively similar above  $3\sigma$ . This strongly suggests that most of the spurious persistence pairs in  $S_N^{2 \times 128}$  do in fact have a persistence ratio lower than  $3\sigma$  and that above that threshold; the remaining persistence pairs have a distribution similar to that in the original  $N$ -body simulation  $S^{128}$ . The persistence pairs of type 0 in  $S_N^{2 \times 128}$  (minima–2-saddle point pairs, blue filled curves) exhibit a slightly different behaviour though, as their number seems to vary more or less in accordance with the persistence threshold in  $S_N^{2 \times 128}$  and  $S_R^{128}$  (blue dotted curve). This number nevertheless always remains higher in  $S_N^{2 \times 128}$  and there are proportionally more high-persistence pairs in  $S_N^{2 \times 128}$  than in  $S_R^{128}$ . This suggests that the number of minima resulting from the physical processes at stake in the void formation is relatively low compared to that due to the Poisson noise, the reason for this being that the cosmological voids’ minima have an intrinsically lower density because of the nature of voids. While the Poisson noise creates spurious minima over a wide range of densities, the voids’ minima only span the lower densities and therefore stretch over comparatively larger scales due to DTFE properties (resolution being inversely proportional to the density). The addition of random particles in  $S_N^{2 \times 128}$  particularly affects the wider regions around minima, therefore increasing their density and lowering the persistence ratio of the corresponding persistence pairs, hence the lack of high-significance pairs of type 0 at  $S(r) > 5\sigma$  (see blue curves) in  $S_N^{2 \times 128}$  compared to  $S^{128}$ . Note, however, that this does not only mean that the physically created persistence pairs are destroyed by the Poisson noise in  $S_N^{2 \times 128}$ , but also mean that they are shifted to lower persistence and that the persistence threshold should not be chosen too high if one wants to retrieve the full DMC

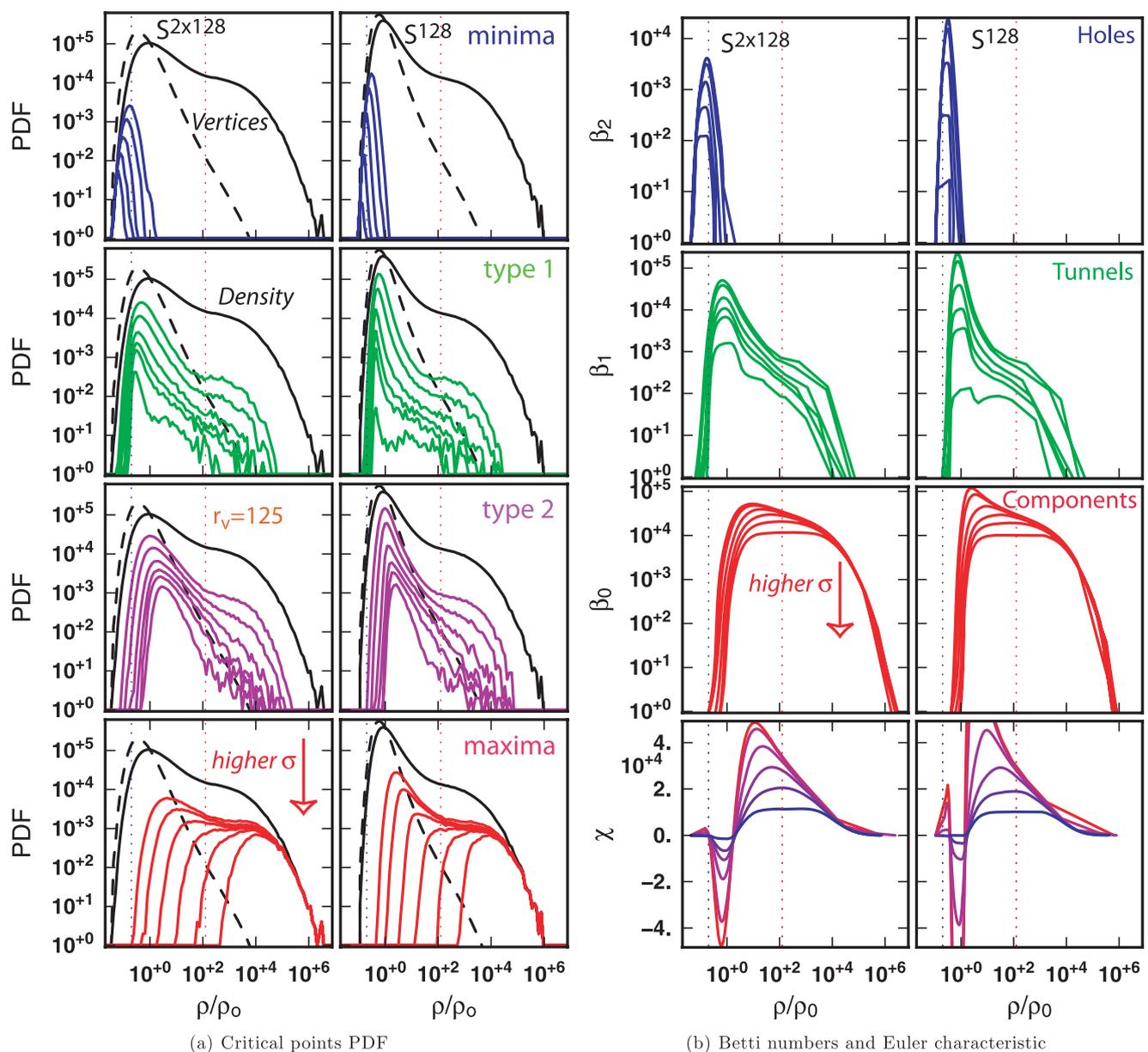
(which is not the case if one is only interested in the filaments). Cosmologically speaking, it basically reflects the fact that clusters and filaments are much stronger topological features in cosmological density fields than in a random particle distribution, whereas the voids, for instance, are not as significantly more pronounced in cosmological density fields as in a random particle distribution.

Two complementary measures of the evolution of the topological properties in  $S^{128}$  and  $S_N^{2 \times 128}$  with the persistence threshold are presented in Fig. 15: the PDF of the critical points in Fig. 15(a) and the Betti numbers and Euler characteristics in Fig. 15(b).

## 4.2 Critical points

Let us consider Fig. 15(a) first. In that figure, the PDF of the density at vertices (i.e. the particles in the studied distribution) is shown by the dark black bold curve and it is striking how the PDF of the critical points tends to follow it, especially at low persistence (outer curves): the more the  $k$ -simplexes at a given density level, the higher the number of detected critical points of order  $k$ . This is an expected result when the Poisson noise dominates as it affects indifferently any scales, but it is not desirable though as the filamentary structure of the cosmic web is an intrinsic property which should not depend on the properties of a particular sampling technique. One would in fact rather expect the PDF of the critical points to follow the PDF of the volume-weighted density, or equivalently as we use the DTFE, of the number of vertices at a given density in the tessellation.<sup>8</sup> The black bold dashed curve traces the *volume-weighted* PDF of the density at vertices. It is clear from Fig. 15(a) that in the case of the minima, 1-saddle-point and 2-saddle-point PDFs, the bias towards higher better sampled densities due to the DTFE is progressively wiped out with the increasing persistence ratio threshold and almost disappears above a significance level threshold of  $\sim 3\sigma$  (see blue, green and purple curves). The PDF of the maxima though (red curves) exhibits an opposite tendency, as their PDF concentrates at higher densities with increasing persistence ratio thresholds. This actually reflects the nature of the distribution of the dark matter over large scales in the universe. In fact, most maxima are expected to be found within gravitationally bound structures undergoing the non-linear regime (i.e. dark matter haloes), which therefore exhibits densities several order of magnitude higher than the average density and with very steep gradients. (Note that this fact also prevents them from being affected by the Poisson noise too much.) Those regions, although numerous, represent only a very small fraction of the total volume, as reflected by the discrepancy between the PDF of the maxima at high-persistence ratio and the volume-weighted PDF of the density. To confirm these hypotheses, we traced in Figs 15(a) and (b) the blue and red vertical dotted lines which mark the characteristic average underdensity of a void in a Einstein–de Sitter model,  $\rho/\rho_0 \leq 0.2$  (see Blumenthal et al. 1992; Sheth & van de Weygaert 2004; Neyrinck 2008), and the typical critical overdensity above which gravitationally bound structures are identified using a FOF algorithm,  $\rho/\rho_0 \geq 125$  (Summers et al. 1995), respectively. While this is not clear at low-persistence thresholds because of the Poisson noise, all maxima (minima) belonging to persistence pairs with the persistence ratio greater than  $\sim 3\sigma$  have densities above (below) those critical thresholds, while the two types of saddle points lie

<sup>8</sup> In the case of the DTFE, the density of a sample particle is defined as the inverse volume of its dual Voronoi cell and the volume it occupies is also the volume of this cell, which implies that the PDF of the volume-weighted density and that of the number of sample particles are identical.



**Figure 15.** Evolution of the topological properties in a  $512^3$ -particle,  $250 h^{-1}$  Mpc dark matter simulation downsampled to  $128^3$  particles,  $S^{128}$ , for increasing persistence levels (left-hand columns in each figure) and in the same distribution with  $128^3$  additional randomly located particles,  $S_N^{2 \times 128}$  (right-hand columns in each figure). In each panel, the persistence selection level ranges from  $0\sigma$  for the outer coloured curve to  $6\sigma$  for the inner curve. Left-hand panel: the PDF of critical points of type 0 (top panel) up to 3 (bottom panel) as a function of their overdensity  $\rho/\rho_0$ . The black curve is the PDF of the vertices in the tessellation, while the dashed curve stands for the (volume-weighted) PDF of the overdensity  $\rho/\rho_0$ . The blue and red vertical dotted lines emphasize the critical level  $r_v = \rho_v/\rho_0 = 0.2$  ( $r_p = \rho_p/\rho_0 = 125$ ) below (above) which a void (a peak) may be considered physically significant. Right-hand panel: from the top to bottom, the Betti numbers,  $\beta_2$ ,  $\beta_1$  and  $\beta_0$ , and the Euler characteristic  $\chi$  of the excursion set with the overdensity greater than  $\rho/\rho_0$ .

within those limits. This means that the detected persistent maxima and minima correspond to physically meaningful objects, which strongly supports the pertinence of using the persistence-based cancellation of a Morse–Smale complex to identify the characteristic components of the cosmic web, such as cosmic voids and filaments.

### 4.3 Discrete topological invariants

The Betti numbers and Euler characteristics represented in Fig. 15(b) are slightly more involved topological analysis tools than the PDF of critical points (see Paper I for a more formal definition of the Betti numbers and a simple example of their computation).

The  $k$ th Betti number  $\beta_k$  counts the number of  $k$ -cycles in excursion sets as a function of the density threshold of the excursion. Within the context of the 3D cosmological matter distribution, there are three Betti numbers that count the number of holes or 2-cycles ( $\beta_2$ ), the number of tunnels or 1-cycles ( $\beta_1$ ) and the number of distinct components or 0-cycles ( $\beta_0$ ) enclosed in the set of points with the density threshold larger than the aforementioned density threshold. As this threshold decreases, new  $k$ -cycles may be created or destroyed, therefore increasing or decreasing the value of the corresponding Betti numbers. The value of the Betti numbers as a function of the density threshold reflects the global topology of the field (i.e. the way it connects as a function of the density threshold)

and it is therefore very instructive to compare the Betti numbers of two distributions to appreciate how similar or distinct they may be from a topological point of view (see also van de Weygaert et al. 2010 for a presentation of the Betti number interpretation in the context of the cosmic web). For that reason, we plotted in Fig. 15(b), from the top to bottom, the values of  $\beta_2$ ,  $\beta_1$ ,  $\beta_0$  and the Euler characteristic  $\chi$  (a topological invariant, computed as the alternate sum of the Betti numbers) as measured in  $S^{128}$  and  $S_N^{2 \times 128}$  (left-hand and right-hand columns, respectively). As noted in Paper I, the notions of persistence pairs and Betti numbers are intimately related: the Betti numbers were readily computed from the persistence pairs, the positive critical point of order  $k + 1$  increasing  $\beta_k$  when it enters the excursion and the negative critical point of order  $k$  decreasing  $\beta_k$ . The distribution  $S_N^{2 \times 128}$  was obtained by adding an equal number of randomly distributed particles to the particles in the  $N$ -body simulations  $S^{128}$ , and the Betti numbers of the two distributions should therefore give some insight into how the topology is affected by the Poisson noise. Note that the presence of the Poisson noise in  $S_N^{2 \times 128}$  affects the PDF of the sampled density by slightly downscaling it (numerous random particles land in large-scale void regions, increasing their densities, while few of them affect the high-density regions, therefore lowering their density contrast, see the black plain curves in Fig. 15). When comparing Betti numbers in the two distributions, one would rather want to know whether the same structures (i.e. void, tunnel, component, etc.) exist in both distributions, though they exist at slightly different densities. It is therefore more important to compare the general shape and amplitude of the Betti number in both distributions than their value at a precise density threshold. Inspecting Fig. 15(b), it is clear that random particles mainly affect the topological properties of the field around the average density  $\rho_0$ , Betti numbers differing by about an order of magnitude in  $S^{128}$  (left-hand panel) and in  $S_N^{2 \times 128}$  (right-hand panel) at a level around  $\rho/\rho_0 = 1$ . The situation largely improves after the cancellation of the lower persistence pairs though and it is striking how the shape and amplitude of the Betti numbers at a level of the persistence ratio of  $3 \sim 4\sigma$  become similar. Note also that  $\beta_0$  is the Betti number that is the least affected by the Poisson noise and for the persistence higher than  $3\sigma$ , the values are almost identical in  $S^{128}$  and  $S_N^{2 \times 128}$ . This means that individual components in the filtration are created and merged in a very similar way independent of the presence of the Poisson noise, which does not affect the filamentary structure of  $S^{128}$ . It is therefore reasonable to trust the filaments detected at persistence levels higher than  $\sim 3\sigma$  as being true topological properties of the underlying distribution. One should none the less remain cautious with the identification of voids and walls. In fact, although the topology of the 1-cycles and 2-cycles seems to be correctly recovered in  $S_N^{2 \times 128}$  at a significance level of  $3 \sim 4\sigma$ , this is not the case anymore at higher levels and one should be careful not to set the threshold too high. In fact, the cosmological voids and walls are more affected by the Poisson noise as they usually live at densities around  $\rho/\rho_0 = 1$  where the influence of the Poisson noise is the maximum and the corresponding persistence pairs have statistically lower persistence ratios than that associated to filaments.

## 5 CONCLUSION

In this paper, we applied DisPerSE (Paper I) to realistic 3D dark matter cosmological simulations and observed redshift catalogues from the SDSS DR7. We showed that DisPerSE traces very well the observed filaments, walls and voids seen both in simulations and in observations. In either setting, filaments are shown to connect on to haloes, outskirts walls, which circumvent voids, as is topologi-

cally required by the Morse theory. Indeed, DisPerSE warrants that all the well-known and extensively studied mathematical properties of the Morse theory are ensured by the construction at the mesh level. As illustrated in Section 3, DisPerSE assumes nothing about the geometry of the survey or its homogeneity and yields a natural (topologically motivated) self-consistent criterion for selecting the significance level of the identified structures. We demonstrated that the extraction is possible even for very sparsely sampled point processes, as a function of the persistence ratio (a measure of the significance of topological connections between critical points), which allows us to account consistently for the shot noise of real surveys. The corresponding recovered cosmic web is also ‘persistent’ in as much as it is robust because it relies on intrinsic topological features of the underlying density field. Hence, we can now trace precisely the locus of filaments, walls and voids from such samples and assess the confidence of the post-processed sets as a function of this threshold, which can be expressed relative to the expected amplitude of the shot noise. DisPerSE has also been shown to be robust as it is able to recover consistent structures even when the sampling of initially well-resolved catalogues is drastically downgraded. Within a cosmic framework, this criterion was shown to level with a FOF structure finder for the identifications of peaks, while DisPerSE also identifies the connected filaments and quantitatively produces on-the-fly the full set of Betti numbers (number of holes, tunnels, connected components, etc.) *directly from the particles*, as a function of the persistence threshold (as these directly follow from the persistence pairs). We investigated the evolution of the critical points, the Betti numbers and the Euler characteristic as a function of the persistence ratio: it illustrates the biases involved in filtering low-persistence ratios. For dark matter simulations, this criterion was shown to be sufficient even if one particle out of two is noise, when the persistence ratio is set to  $3\sigma$  or more. We also applied this procedure to the localization of a specific filamentary configuration and observed an ‘optically faint’ cluster at a galaxy filament junction, identified in the SDSS catalogue. An X-ray counterpart could indeed be observed (Kawahara et al. 2011) by the X-ray satellite *Suzaku*. All these results are very encouraging for future investigations using DisPerSE, for searching galaxy clusters, galaxy groups and missing baryons of the universe in particular, and for the study of the LSS in general.

## ACKNOWLEDGEMENTS

The authors thank T. Nishimichi for his help in dealing with the SDSS galaxy catalogue, H. Yoshitake for his help with the X-ray analysis and TS thanks Y. Suto for his constant help and support.

This work was made possible through an extensive usage of the YORICK programming language by D. Munro (available at <http://yorick.sourceforge.net/>) and also CGAL, the Computational Geometry Algorithms Library (<http://www.cgal.org/>), to compute the Delaunay tessellations.

Funding for the SDSS has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society and the Higher Education Funding Council for England. The SDSS website is <http://www.sdss.org/>. TS gratefully acknowledges support from the Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowship for Foreign Researchers award P08324. HK is supported by a JSPS (Japan Society for Promotion of Science) Grant-in-Aid for Science Fellows. CP acknowledges support from a Leverhulme Visiting Professorship at

the Astrophysics Department of the University of Oxford. This work is also supported by a Grant-in-Aid for Scientific Research from the JSPS and from the Japanese Ministry of Education, Culture, Sports, Science and Technology (No. 22-5467).

## REFERENCES

- Abazajian K. N. et al., 2009, *ApJS*, 182, 543
- Aracil B., Petitjean P., Pichon C., Bergeron J., 2004, *A&A*, 419, 811
- Aragón-Calvo M. A., van de Weygaert R., Araya-Melo P. A., Platen E., Szalay A. S., 2010a, *MNRAS*, 404, L89
- Aragón-Calvo M. A., Platen E., van de Weygaert R., Szalay A. S., 2010b, *ApJ*, 723, 364
- Blanton M. R. et al., 2005, *AJ*, 129, 2562
- Blumenthal G. R., da Costa L. N., Goldwirth D. S., Lecar M., Piran T., 1992, *ApJ*, 388, 234
- Buote D. A., Zappacosta L., Fang T., Humphrey P. J., Gastaldello F., Tagliaferri G., 2009, *ApJ*, 695, 1351
- Cen R., Ostriker J. P., 1999, *ApJ*, 514, 1
- Cohen-Steiner D., Edelsbrunner H., Harer J., 2007, *Discrete Comput. Geom.*, 37, 103
- Danforth C. W., Stocke J. T., Shull J. M., 2010, *ApJ*, 710, 613
- de Lapparent V., Geller M. J., Huchra J. P., 1986, *ApJ*, 302, L1
- Edelsbrunner H., Letscher D., Zomorodian A., 2000, in 41st Ann. Symp. Foundations of Computer Science, Topological Persistence and Simplification. IEEE Comput. Soc. Press, Los Alamitos, CA, p. 454
- Edelsbrunner H., Letscher D., Zomorodian A., 2002, *Discrete Comput. Geom.*, 28, 511
- Efstathiou G., Davis M., White S. D. M., Frenk C. S., 1985, *ApJS*, 57, 241
- Fang T., Bryan G. L., Canizares C. R., 2002, *ApJ*, 564, 604
- Fang T., Buote D. A., Humphrey P. J., Canizares C. R., Zappacosta L., Maiolino R., Tagliaferri G., Gastaldello F., 2010, *ApJ*, 714, 1715
- Forero-Romero J. E., Hoffman Y., Gottlöber S., Klypin A., Yepes G., 2009, *MNRAS*, 396, 1815
- Forman R., 2002, *Sém. Lothar. Combin. Art.*, 2, 21
- Fukugita M., Hogan C. J., Peebles P. J. E., 1998, *ApJ*, 503, 518
- Gay C., Pichon C., Le Borgne D., Teyssier R., Sousbie T., Devriendt J., 2010, *MNRAS*, 404, 1801
- Hahn O., Porciani C., Carollo C. M., Dekel A., 2007, *MNRAS*, 375, 489
- Jost J., 2008, *Riemannian Geometry and Geometric Analysis*, 5th edn. Universitext, Springer-Verlag, Berlin
- Kawahara H., Yoshikawa K., Sasaki S., Suto Y., Kawai N., Mitsuda K., Ohashi T., Yamasaki N. Y., 2006, *PASJ*, 58, 657
- Kawahara H., Yoshitake H., Nishimichi T., Sousbie T., 2011, *ApJ*, 727, L38
- Milnor J., 1963, *Annals of Mathematics Studies*, No. 51. Princeton Univ. Press, Princeton, NJ
- Mitsuda K. et al., 2007, *PASJ*, 59, 1
- Neyrinck M. C., 2008, *MNRAS*, 386, 2101
- Noh Y., Cohn J. D., 2011, *MNRAS*, in press (doi:10.1111/j.1365-2966.2010.18137.x) (arXiv:1011.1000)
- Novikov D., Colombi S., Doré O., 2006, *MNRAS*, 366, 1201
- Ohashi T. et al., 2006, in Turner M. J. L., Hasinger G., SPIE Conf. Ser. Vol. 6266, *Space Telescopes and Instrumentation II: Ultraviolet to Gamma Ray*, SPIE, Bellingham, p. 62660G
- Platen E., van de Weygaert R., Jones B. J. T., 2007, *MNRAS*, 380, 551
- Pogosyan D., Bond J. R., Kofman L., Wadsley J., 1996, *BAAS*, 28, 1289
- Pogosyan D., Pichon C., Gay C., Prunet S., Cardoso J. F., Sousbie T., Colombi S., 2009, *MNRAS*, 396, 635
- Schaap W. E., van de Weygaert R., 2000, *A&A*, 363, L29
- Sheth R. K., van de Weygaert R., 2004, *MNRAS*, 350, 517
- Sousbie T., 2011, *MNRAS*, in press (doi:10.1111/j.1365-2966.2011.18394.x) (Paper I, this issue)
- Sousbie T., Pichon C., Courtois H., Colombi S., Novikov D., 2008a, *ApJ*, 672, L1
- Sousbie T., Pichon C., Colombi S., Novikov D., Pogosyan D., 2008b, *MNRAS*, 383, 1655
- Sousbie T., Colombi S., Pichon C., 2009, *MNRAS*, 393, 457
- Stoica R. S., Martínez V. J., Mateu J., Saar E., 2005, *A&A*, 434, 423
- Stoica R. S., Martínez V. J., Saar E., 2007, *J. R. Stat. Soc. C (Appl. Stat.)*, 56, 459
- Stoica R. S., Martínez V. J., Saar E., 2010, *A&A*, 510, A38
- Summers F. J., Davis M., Evrard A. E., 1995, *ApJ*, 454, 1
- Tago E., Saar E., Tempel E., Einasto J., Einasto M., Nurmi P., Heinämäki P., 2010, *A&A*, 514, A102
- Tripp T. M., Savage B. D., Jenkins E. B., 2000, *ApJ*, 534, L1
- van de Weygaert R., 2002, in Plionis M., Cotsakis S., eds, *Astrophys. Space Sci. Libr.*, Vol. 276, *Modern Theoretical and Observational Cosmology*. Kluwer, Dordrecht, p. 119
- van de Weygaert R., Schaap W., 2009, in Martínez V. J., Saar E., Gonzales E. M., Pons-Borderia M. J., eds, *The Cosmic Web: Geometric Analysis*. Springer-Verlag, Berlin, p. 291
- van de Weygaert R., Platen E., Vegter G., Eldering B., Kruithof N., 2010, *Int. Symp. Voronoi Diagr. Sci. Eng.*, 0, 224
- Yoshikawa K., Yamasaki N. Y., Suto Y., Ohashi T., Mitsuda K., Tawara Y., Furuzawa A., 2003, *PASJ*, 55, 879

## APPENDIX A: VORONOI KINEMATIC MODELS AS A TEST CASE FOR MORSE–SMALE COMPLEX IDENTIFICATION

Assessing the quality of the identification of the filamentary structures requires the previous knowledge of their location. One is therefore tempted to try building an artificial distribution from an ad hoc Morse–Smale complex, preferentially with properties similar to that of a cosmological density field. This method was adopted in Aragón-Calvo et al. (2010b), where the authors use a Voronoi kinematic model (van de Weygaert 2002). The principle of the Voronoi kinematic model is to identify the voids, walls, filaments and clusters to the cells, faces, edges and vertices of the Voronoi tessellation. In practice, randomly distributed particles are moved away from the nuclei of the Voronoi cells following a universal expansion rate and their displacement being constrained to the faces, edges and finally vertices as they reach them. This results in a distribution of particles where each is said to be a void, wall, filament or cluster particle depending on whether they belong to a cell, face, edge or vertex of a Voronoi cell when the simulation is stopped.

We argue that using such a model to quantify the quality of the Morse–Smale complex identification is not as relevant as one would think, mainly because it is too idealized, topologically speaking. In fact, it is a built-in property of the Voronoi kinematic models that all the cosmological structures overlap neatly: maxima (i.e. voronoi vertices) are located at the intersection of filaments (i.e. Voronoi segments) that always intersect with a suitable angle, those filaments are themselves by definition located at the intersection of at least three voids (i.e. voronoi cells), and each pair of a neighbouring void has exactly one Voronoi face in common, neatly defining the walls. As was shown in Paper I, density functions extracted from actual data sets are in fact quite different, as they do not comply so easily to Morse conditions, in particular when measured from cosmological simulations or observational galaxy catalogues. In that case, and as clearly shown in Paper I (see appendix 1), filaments may (and actually often do) merge before reaching a maximum, two apparently neighbouring voids (down to the resolution limits) do not necessarily share a common face, and filaments are not necessarily at the intersection of at least three voids (once again, down to the resolution limit). The nature of the Voronoi kinematic model is therefore such that it avoids all the difficulty in identifying the Morse–Smale complex of realistic data sets. It might be possible to build more sophisticated Voronoi Models that would, for instance, mimic the structure mergers that occur along the course of the

evolution of the large-scale matter distribution in the Universe, but this is clearly out of the scope of this paper.

## APPENDIX B: THE ALGORITHM IN A NUTSHELL

Let us summarize the basis of the algorithm introduced in Paper I, while connecting the corresponding jargon to the more intuitive language of astrophysical data processing. For the purpose of the skeleton extraction, a simplicial complex (the tessellation) is computed from a discrete distribution (galaxy catalogue,  $N$ -body simulation, etc.) using a Delaunay tessellation. A density  $\rho$  is assigned to each galaxy using the DTFE [roughly speaking, the density at a vertex is proportional to the inverse volume of its dual Voronoi cell, see Schaap & van de Weygaert (2000) and van de Weygaert & Schaap (2009)]. A discrete Morse function (a relabelling of all elements of the tessellation) is then defined by attributing a properly chosen value to each simplex in the complex (i.e. the segments, facets and tetrahedron of the tessellation). From this discrete function, we then compute the discrete gradient and deduce the corresponding DMC (Forman 2002). The DMC (the set of critical points connected by arcs, quads, crystals, etc.) is used as the link between the topological and geometrical properties of the density field. Its critical points together with their ascending and descending manifolds (the ‘critical’ sets) are identified to the peaks, filaments, walls and voids of the density field. The DMC is then filtered using the persistence theory. For that purpose, we consider the filtration (the discrete counterpart of the density-sorted-level sets) of the tessellation according to the values of the discrete Morse function and use it to compute persistence pairs of critical points (pairs of critical points that create and destroy a given topological feature as the filtration grows). The DMC is simplified by cancelling the pairs that are likely to be generated by noise. This is achieved by computing the PDF of the persistence ratio (i.e. the ratio of the densities at the connected pair) of all types of pairs in scale-invariant Gaussian random fields and cancelling the pairs with a persistence ratio whose probability is lower than a certain level.

## APPENDIX C: TERMINOLOGY

*Arc.* An arc is a 1-cell: an integral line (or a V-path in the discrete theory) whose origin and destinations are critical points. The arcs of the Morse–Smale complex connect two critical points of order difference 1 (i.e. in 2D, a minimum and a saddle point or a maximum and a saddle point).

*n-cell.* A  $n$ -cell is a region of space of dimension  $n$  such that all the integral lines in the  $n$ -cell have a common origin and destination. The  $n$ -cells basically partition space into regions of uniform gradient flow.

*Coface.* A coface of a  $k$ -simplex  $\alpha_k$  is any  $p$ -simplex  $\beta_p$ , with  $p \geq k$ , such that  $\alpha_k$  is a face of  $\beta_p$ . In 3D, the cofaces of a segment (i.e. a 1-simplex) are any triangle or tetrahedron (i.e. 2- or 3-simplex) whose set of summits (i.e. vertices) contains the two vertices at the extremities of the segment, as well as the segment itself.

*Cofacet.* A cofacet of a  $k$ -simplex  $\alpha_k$  is a coface  $\beta_{k+1}$  of  $\alpha_k$  with dimension  $k + 1$ . Equivalently,  $\alpha_k$  is a facet of  $\beta_{k+1}$ .

*Critical point of order  $k$ .* For a smooth function  $f$ , a critical point of order  $k$  is a point such that the gradient of  $f$  is null and the Hessian (matrix of second derivatives) has exactly  $k$  negative eigenvalues. In 2D, a minimum, saddle point and maximum are critical points of orders 0, 1 and 3, respectively.

*Critical  $k$ -simplex.* A critical  $k$ -simplex is the equivalent in the discrete Morse theory of the critical point of order  $k$  in its smooth counterpart. Note that in 2D, the equivalent of a minimum is a critical vertex (0-simplex), a saddle point is a critical segment (1-simplex) and a maximum is a critical triangle (2-simplex).

*Crystal.* A crystal is a 3-cell: a 3D region typically delimited by six quads and 12 arcs, within which all the integral lines (or V-paths in the discrete case) have identical origin and destinations.

*k-cycle.* A  $k$ -cycle in a simplicial complex corresponds to a  $k$ -dimensional topological feature. In 3D, 0-cycles correspond to independent components, 1-cycles to loops and 2-cycles to shells.

*Discrete gradient.* A discrete gradient of a discrete Morse–Smale function  $f$  defined over a simplicial complex  $K$ -pair simplexes of  $K$ . Within a gradient pair, the simplex with a lower value is called the tail and the other the head, and any unpaired simplex is critical.

*DMC.* The DMC is the equivalent of the Morse–Smale complex applied to simplicial complexes.

*Discrete Morse–Smale function.* A discrete Morse–Smale function  $f$  defined over a simplicial complex  $K$  associates a real value  $f(\sigma_k)$  to each simplex  $\sigma_k \in K$ .

*Excursion set.* An excursion set of a function  $\rho(x)$  is the set of points for which  $\rho(x) \geq \rho_0$  (see also the sublevel set).

*Face.* A face of a  $k$ -simplex  $\alpha_k$  is any  $p$ -simplex  $\beta_p$  with  $p \leq k$ , such that all vertices of  $\beta_p$  are also vertices of  $\alpha_k$ . In 3D, the faces of a 3-simplex (i.e. a tetrahedron) are the tetrahedron itself, the four triangles that form its boundaries, the six segments that form its edges and its four summits (i.e. vertices).

*Facet.* A facet of a  $k$ -simplex  $\alpha_k$  is a face  $\beta_{k-1}$  of  $\alpha_k$  with dimension  $k - 1$ . The facets of a 3-simplex (i.e. a tetrahedron) are the four triangles (i.e. 2-simplexes) that form its boundaries.

*Filtration.* A filtration of a simplicial complex  $K$  is a growing sequence of subcomplexes  $K_i$  of  $K$ , such that each  $K_i$  is also a simplicial complex. If the different  $K_i$  are defined by a discrete function  $F_\rho$  as the set of simplexes of  $K$  with values  $F_\rho(\sigma)$  less than or equal to a given threshold, then a filtration can be thought as the discrete equivalent of a sequence of growing sublevel sets of a smooth function.

*Gradient pair/arrow.* A gradient pair or arrow is a set of two simplexes, one being the facet of the other, and such that they are paired within a discrete gradient. Within a gradient pair, the simplex with the lower value is called the tail and the other the head.

*Integral line.* An integral line of a scalar function  $\rho(x)$  is a curve whose tangent vector agrees with the gradient of  $\rho(x)$ .

*Level set/sublevel set.* A level set, also called isocontour, of a function  $\rho(x)$  at level  $\rho_0$  is the set of points such that  $\rho(x) = \rho_0$ . The corresponding sublevel set is the set of points such that  $\rho(x) \leq \rho_0$ .

*Ascending/descending  $p$ -manifold.* Within a space of dimension  $d$ , an ascending  $p$ -manifold is the set of points from which, following minus the gradient, one reaches a given critical point of order  $d - p$ . A descending  $p$ -manifold is the set of points from which, following the gradient, one reaches a given critical point of order  $p$ . For instance, ascending 1-manifolds in 3D can be associated to the filaments and ascending 3-manifolds describes the voids.

*Morse function.* A Morse function is a continuous, twice-differentiable smooth function whose critical points are non-degenerate. In particular, the eigenvalues of the Hessian matrix (i.e. the matrix of the second derivatives) must be non-null.

*Morse complex.* The Morse complex of a Morse function is the set of its ascending (or descending) manifolds.

*Morse–Smale function.* A Morse–Smale function is a Morse function whose ascending and descending manifolds intersect

*transversely*. This means that there exists no point where an ascending and a descending manifold may be tangent.

*Morse–Smale complex*. The Morse–Smale complex is the intersection of the ascending and descending manifolds of a Morse–Smale function. One can think of the Morse–Smale complex as a network of critical points connected by  $n$ -cells, defining a notion of hierarchy and neighbourhood among them. In particular, the geometry of the arcs (i.e. 1-cells) is determined by the critical integral lines (i.e. integral lines that join critical points) and the order-2 critical points connected by an arc may only differ by 1.

*Peak/void patch*. In 3D, a peak patch is a descending 3-manifold (i.e. the region of space from which, following the gradient, one reaches a given maximum) and a void patch an ascending 3-manifold (i.e. the region of space from which, following minus the gradient, one reaches a given minimum).

*Persistence*. The persistence of a persistence pair (or equivalently of the corresponding  $k$ -cycle it creates and destroys) is defined as the difference between the values of the two critical points (or critical simplexes in the discrete case) in the pair. It basically represents its lifetime within the evolving sublevel sets (or filtration in the discrete case).

*Persistence pair*. In the smooth context of a function  $\rho$ , the persistence pair critical points  $P_a$  and  $P_b$  of  $\rho$  that, respectively, create and destroy a topological feature (or  $k$ -cycle) in the sublevel sets of  $\rho$ , at levels  $\rho(P_a)$  and  $\rho(P_b)$ . In the discrete case of a simplicial complex  $K$ , a persistence pair is a pair of critical simplexes  $\sigma_a$  and  $\sigma_b$  of a given discrete function  $F_\rho(\sigma)$ , such that  $\sigma_a$  creates a  $k$ -cycle (i.e. topological feature) when it enters the filtration of  $K$  according to  $F_\rho$ , and  $\sigma_b$  destroys it when it enters.

*Persistence ratio*. The persistence ratio of a persistence pair (or equivalently of the corresponding  $k$ -cycle it creates and destroys) is the ratio of the value of the two critical points (or critical simplexes in the discrete case) in the pair. The persistence ratio is preferred to the regular persistence in the case of strictly positive functions such as the density field of matter on large scales in the universe.

*Quad*. A quad is a 2-cell: a 2D region delimited by four arcs within which all the integral lines (or V-paths in the discrete case) have identical origin and destinations.

*$k$ -simplex*. A  $k$ -simplex is the  $k$ -dimensional analogue of a triangle: the simplest geometrical object with  $k + 1$  summits, called vertices. It is the building block of simplicial complexes.

*Simplicial complex*. A simplicial complex  $K$  is a set of simplexes such that if a  $k$ -simplex  $\alpha_k$  belongs to  $K$ , then all its faces also belong to  $K$ . Moreover, the intersection of two simplexes in  $K$  must be a simplex that also belongs to  $K$ .

*Vertex*. A vertex is a 0-simplex or simply a point.

*V-path*. A V-path is the discrete equivalent of an integral line: it is a set of simplexes linked by discrete gradient arrows and the face–coface relation. Tracing a V-path consists in intuitively following the direction of the gradient pairs of a discrete gradient from a critical simplex to another.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.