Appendix C

Cosmic structures identification and classification algorithms

Contents

"Whenever a theory appears to you as the only possible one, take this as a sign that you have neither understood the theory nor the problem which it was intended to solve."
— Karl Popper (1972), *Objective Knowledge: An Evolutionary Approach*

Abstract

This appendix discusses methods for identifying and classifying structures in the cosmic web. As many approaches exist (see the introduction of chapter 9), in the following we only focus on the algorithms used in this thesis: the VIDE toolkit for the identification of static voids (section C.1), and the T-web approach for dissecting the dynamic cosmic web into clusters, filaments, sheets, and voids (section C.2).

C.1 VIDE: the Void IDentification and Examination toolkit

This section describes VIDE, the Void IDentification and Examination toolkit. It is a static void finder operating on density fields, used in chapter 8 of this thesis. The details behind VIDE are described in its accompanying paper, Sutter *et al.* (2015b), and its website http://www.cosmicvoids.net/. VIDE is based on ZOBOV (ZOnes Bordering On Voidness, Neyrinck, 2008) for the void finding part (sections C.1.1 and C.1.2), and includes a set of additional features for pre- and post-processing void catalogs (section C.1.3).

C.1.1 Voronoi Tessellation Density Estimation

The algorithm begins by building a Voronoi tessellation of the tracer particle population (Schaap & van de Weygaert, 2000; Schaap, 2007). This provides a density field estimator (the Voronoi Tessellation Field Estimator, VTFE) based on the underlying particle positions. The VTFE (along with its dual, the Delaunay

Tessellation Field Estimator, DTFE) is a local density estimate that is especially suitable for astronomical data (van de Weygaert & Schaap, 2009; Cautun & van de Weygaert, 2011).

The Voronoi tessellation is a partitioning of space into cells around each particle. For each particle i, the corresponding Voronoi cell is the region consisting of all points closer to that particle than to any other. The density estimate at particle i is $1/V(i)$, where $V(i)$ is the volume of the Voronoi cell around particle i. It is further assumed constant density across the volume of each Voronoi cell, which effectively sets a smoothing scale for the continuous density field.

Finally, the Voronoi tessellation also provides the adjacency measurement for each particle i (i.e. the set of particles whose Voronoi cells have a common boundary with i's cell), which ZOBOV uses in the next step.

C.1.2 The watershed algorithm

ZOBOV then uses the watershed transform (Platen, van de Weygaert & Jones, 2007) to group Voronoi cells into zones and subsequently voids. Minima (also called cores or basins) are first identified as particles with lower density than any of their Voronoi neighbors. Then, the algorithm merges nearby Voronoi cells into zones (the set of cells for which density flows downward into the zone's core). Finally, the watershed transform groups adjacent zones into voids by finding minimum-density barriers between them and joining zones together. This can be thought of, for each zone z, as setting the "water level" to its minimum density and raising it gradually. Water may flow along lines joining adjacent Voronoi zones, adding them to the void defined around zone z. The process is stopped when water flows into a deeper zone (one with a lower core than z) or if z is the deepest "parent" void, when water floods the whole field. The void corresponding to zone z is defined as the set of zones filled with water just before this happens, and its boundary is the ridgeline which retains the flow of water. As can be understood from this description, the watershed transform naturally builds a nested hierarchy of voids (Lavaux & Wandelt, 2012; Bos *et al.*, 2012).

ZOBOV imposes a density-based criterion within the void finding operation: adjacent zones are only added to a void if the density of the wall between them is less than 0.2 times the mean particle density (Platen, van de Weygaert & Jones, 2007; see Blumenthal *et al.*, 1992; Sheth & van de Weygaert, 2004 for the role of the corresponding $\delta = -0.8$ underdensity). This density threshold prevents voids from expanding deeply into overdense structures and limits the depth of the void hierarchy (Neyrinck, 2008). By default, VIDE reports every identified basin as a void (regardless of the density of the initial zone), but facilities exist for filtering the void catalogs based on various criteria (Sutter *et al.*, 2015b).

C.1.3 Processing and analysis of void catalogs

The VIDE toolkit provides routines for performing many analysis tasks, such as manipulating, filtering, and comparing void catalogs, plotting void properties, stacking, computing clustering statistics and fitting density profiles (Sutter *et al.*, 2015b). In this section, we briefly describe the details behind the three void statistics used in chapter 8: number functions, ellipticity distributions, and density profiles.

C.1.3.1 Number functions

The effective radius of a void is defined as

$$R_\mathrm{v} \equiv \left(\frac{3}{4\pi} V \right)^{1/3}, \tag{C.1}$$

where V is the total volume of the Voronoi cells that make up the void. From this definition, voids with effective radius smaller than $\bar{n}^{-1/3}$, where $\bar{n}$ is the mean number density of tracers, are excluded to prevent the effects of shot noise.

Based on this definition, VIDE includes a built-in plotting routine for the cumulative number functions of multiple void catalogs on a logarithmic scale (see figure 8.3).

C.1.3.2 Ellipticity distributions

For each void in the catalog, VIDE also reports the volume-weighted center of all Voronoi cells in the void, or macrocenter:

$$\mathbf{x}_\mathrm{v} \equiv \frac{1}{\sum_i V_i} \sum_i \mathbf{x}_i V_i, \tag{C.2}$$

where $\mathbf{x}_i$ and V_i are the positions and Voronoi volumes of each tracer particle i, respectively.

Void shapes are computed from void member particles by constructing the inertia tensor:

$$M_{xx} = \sum_{i=1}^{N_\mathrm{p}} \left(y_i^2 + z_i^2\right), \tag{C.3}$$

$$M_{xy} = -\sum_{i=1}^{N_\mathrm{p}} x_i y_i, \tag{C.4}$$

where N_p is the number of particles in the void, and (x_i, y_i, z_i) is the set of coordinates of particle i relative to the void macrocenter. The other components of the inertia tensor are obtained by cyclic permutation of coordinates. The eigenstructure of the inertia tensor gives the ellipticity of the void:

$$\varepsilon = 1 - \left(\frac{J_1}{J_3}\right)^{1/4}, \tag{C.5}$$

where J_1 and J_3 are the smallest and the largest eigenvalues of the inertia tensor, respectively. The ellipticity distribution of voids as a function of their effective radius follows from this definition (see figure 8.4).

C.1.4 Radial density profiles

VIDE contains a routine to construct three-dimensional stacks of voids, where void macrocenters are superposed and particle positions are shifted to be expressed as relative to the stack center. This routine builds stacks of voids whose effective radius is in some given range. From each of these three-dimensional stacks, VIDE builds a spherically-averaged one-dimensional profile.

This is used in particular for building radial density profiles of voids at a given size (see figure 8.5).

C.2 The T-web

This section describes the "T-web", a dynamic web classifier which dissects the entire large-scale structure into different structure types: voids, sheets, filaments, and clusters. It is used in section 2.3, chapters 9 and 10 of this thesis.

C.2.1 The tidal tensor

We start here from the Vlasov-Poisson system in Eulerian coordinates, equations (1.72) and (1.75). It is always possible to rescale the cosmological gravitational potential by defining $\tilde{\Phi} \equiv \Phi/(4\pi \mathrm{G} a^2 \bar{\rho})$ in such a way that $\tilde{\Phi}$ obeys a reduced Poisson equation,

$$\Delta\tilde{\Phi}(\mathbf{x}) = \delta(\mathbf{x}). \tag{C.6}$$

In this context, we define the *tidal tensor* $\mathcal{T}$ as the Hessian $\mathrm{H}(\tilde{\Phi})$ of the rescaled gravitational potential $\tilde{\Phi}$,

$$\mathcal{T}_{ij} \equiv \mathrm{H}(\tilde{\Phi})_{ij} = \frac{\partial^2 \tilde{\Phi}}{\partial \mathbf{x}_i \partial \mathbf{x}_j}. \tag{C.7}$$

With this definition, the left-hand side of equation (C.6) can be seen as the application of the Laplace-Beltrami operator $\mathcal{LB}$ (or tensor Laplacian), trace of the Hessian, to $\tilde{\Phi}$:

$$\mathcal{LB}(\tilde{\Phi}) \equiv \mathrm{tr}(\mathrm{H}(\tilde{\Phi})) = \Delta\tilde{\Phi}. \tag{C.8}$$

Let us denote by $\mu_1(\mathbf{x}) \leq \mu_2(\mathbf{x}) \leq \mu_3(\mathbf{x})$ the three local eigenvalues of the tidal tensor.[1] They are dimensionless and real (since $\mathcal{T}$ is symmetric). We have $\mathrm{tr}(\mathcal{T})(\mathbf{x}) = \mu_1(\mathbf{x}) + \mu_2(\mathbf{x}) + \mu_3(\mathbf{x})$, and the reduced Poisson equation can therefore be seen as a decomposition of the Eulerian density contrast field, in the sense that it reads

$$\mu_1(\mathbf{x}) + \mu_2(\mathbf{x}) + \mu_3(\mathbf{x}) = \delta(\mathbf{x}). \tag{C.9}$$

[1] These eigenvalues are often noted λ_i in the literature. We changed the notation in this thesis to avoid the confusion with the Zel'dovich formalism (see sections 1.5.2 and C.2.2).

At this point, it is useful to introduce some notations commonly found in the literature to characterize the tidal field. Given equation (C.9), the eigenvalues of the tidal tensor define an ellipsoid with semi-axes (e.g. Peacock & Heavens, 1985)

$$a_i(\mathbf{x}) \equiv \sqrt{\frac{\delta(\mathbf{x})}{\mu_i(\mathbf{x})}}. \tag{C.10}$$

The triaxiality parameters are defined by Bardeen *et al.* (1986) in terms of the eigenvalues as

$$\varepsilon(\mathbf{x}) = \frac{\mu_1(\mathbf{x}) - \mu_3(\mathbf{x})}{2\delta(\mathbf{x})} \quad \text{and} \quad p(\mathbf{x}) = \frac{\mu_1(\mathbf{x}) - 2\mu_2(\mathbf{x}) + \mu_3(\mathbf{x})}{2\delta(\mathbf{x})}. \tag{C.11}$$

ε is called the ellipticity (in the $\mu_1 - \mu_3$ plane) and p the prolateness (or oblateness). If $-\varepsilon \leq p \leq 0$ then the ellipsoid is prolate-like, and if $0 \leq p \leq \varepsilon$ it is oblate-like. The limiting cases are $p = -\varepsilon$ for prolate spheroids and $p = \varepsilon$ for oblate spheroids.

C.2.2 Analogy with the Zel'dovich formalism

The above equations have a strong similarity with that of the Zel'dovich formalism. Indeed, we have seen that the first Lagrangian potential $\phi^{(1)}$, defined by $\mathbf{\Psi}^{(1)}(\mathbf{q}, \tau) = -D_1(\tau)\nabla_\mathbf{q}\phi^{(1)}(\mathbf{q})$, satisfies a reduced Poisson equation (equation (1.134)),

$$\Delta_\mathbf{q}\phi^{(1)}(\mathbf{q}) = \delta(\mathbf{q}). \tag{C.12}$$

As discussed in section 1.5.2, the shear of the displacement $\mathcal{R} \equiv \partial\mathbf{\Psi}/\partial\mathbf{q}$ verifies

$$\mathcal{R}_{ij} = -D_1(\tau)\mathrm{H}(\phi^{(1)})_{ij} = -D_1(\tau)\frac{\partial^2\phi^{(1)}}{\partial\mathbf{q}_i\partial\mathbf{q}_j}. \tag{C.13}$$

The local eigenvalues of Hessian of the first Lagrangian potential, $\lambda_1(\mathbf{q}) \leq \lambda_2(\mathbf{q}) \leq \lambda_3(\mathbf{q})$, permit to rewrite the reduced Poisson equation as a decomposition of the initial density contrast,

$$\lambda_1(\mathbf{q}) + \lambda_2(\mathbf{q}) + \lambda_3(\mathbf{q}) = \delta(\mathbf{q}). \tag{C.14}$$

C.2.3 The T-web: original procedure

In analogy with the Zel'dovich "pancake" theory, where the sign of the λ_i permit an interpretation of what happens at shell-crossing in the ZA in terms of structure types (see section 1.5.2), Hahn *et al.* (2007a) proposed to classify structures using the sign of the μ_i. Namely, a void point corresponds to no positive eigenvalue, a sheet to one, a filament to two, and a cluster to three positive eigenvalues (see table C.1).

Structure type	Rule
Void	$\mu_1, \mu_2, \mu_3 < 0$
Sheet	$\mu_1, \mu_2 < 0$ and $\mu_3 > 0$
Filament	$\mu_1 < 0$ and $\mu_2, \mu_3 > 0$
Cluster	$\mu_1, \mu_2, \mu_3 > 0$

Table C.1: Rules for classification of structure types according to the T-web procedure (Hahn *et al.*, 2007a).

The interpretation of this rule is straightforward, as the sign of an eigenvalue at a given position defines whether the gravitational force in the direction of the corresponding eigenvector is contracting (positive eigenvalues) or expanding (negative eigenvalues). Thus, the signature of the tidal tensor characterizes the number of axes along which there is gravitational expansion or collapse. This procedure is sometimes called the "T-web", in reference to the tidal tensor.

In Hahn *et al.* (2007a), an interpretation of the above rule in terms of the orbit stability of test particles is also discussed. The equation of motion in comoving coordinates and in conformal time reads (see equation (1.74))

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}\tau} = -ma\nabla\Phi \quad \text{with} \quad \mathbf{p} = ma\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\tau} \tag{C.15}$$

The local extrema of the gravitational potential (i.e. points $\bar{\mathbf{x}}$ such that $\nabla\Phi(\bar{\mathbf{x}}) = 0$) are fixed points of the test particle equation of motion. These can be, for example, the center of mass of halos. In their neighborhood, we can linearize the equation of motion at the points $\bar{\mathbf{x}}$, which yields the linear system

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left(ma\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\tau}\right) \approx -ma\,\nabla^2\Phi(\bar{\mathbf{x}}) \cdot (\mathbf{x} - \bar{\mathbf{x}})\,, \tag{C.16}$$

or, in terms of coordinates,

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left(ma\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}\tau}\right) \approx -ma\sum_j \frac{\partial^2\Phi}{\partial\mathbf{x}_i\partial\mathbf{x}_j}(\bar{\mathbf{x}})\,(\mathbf{x}_j - \bar{\mathbf{x}}_j) \propto -ma\sum_j \mathcal{T}_{ij}(\bar{\mathbf{x}})\,(\mathbf{x}_j - \bar{\mathbf{x}}_j)\,. \tag{C.17}$$

This equation means that the linear dynamics near local extrema of the gravitational potential is fully governed by the tidal field. The number of positive eigenvalues is equivalent to the dimension of the stable manifold at the fixed points:

- voids are regions of space where the orbits of test particles are unstable (no positive eigenvalue);
- sheets correspond to one-dimensional stable manifolds (one positive, two negative eigenvalues);
- filaments correspond to two-dimensional stable manifolds (two positive, one negative eigenvalues);
- clusters are attractive fixed points (three positive eigenvalues).

Dropping the assumption of local extrema of the gravitational potential introduces a constant acceleration term to the linearized equation of motion. This zeroth-order effect can be ignored by changing to free-falling coordinates. The behavior introduced by the first-order term, representing the tidal deformation of orbits, and thus the web-type classification, remain unchanged.

C.2.4 Extensions of the T-web

C.2.4.1 Varying threshold

Several extensions of this classification procedure exist. Forero-Romero *et al.* (2009) pointed out that rather than using a threshold value μ_{th} of zero, different positive values can be used. The corresponding set of rules is given by table C.2.

Structure type	Rule
Void	$\mu_1, \mu_2, \mu_3 < \mu_{\mathrm{th}}$
Sheet	$\mu_1, \mu_2 < \mu_{\mathrm{th}}$ and $\mu_3 > \mu_{\mathrm{th}}$
Filament	$\mu_1 < \mu_{\mathrm{th}}$ and $\mu_2, \mu_3 > \mu_{\mathrm{th}}$
Cluster	$\mu_1, \mu_2, \mu_3 > \mu_{\mathrm{th}}$

Table C.2: Rules for classification of structure types according to the extended T-web procedure with varying threshold (Forero-Romero *et al.*, 2009).

This introduces a new free parameter, which *a priori* can take any value. However, Forero-Romero *et al.* (2009) argued that a natural threshold can be roughly estimated by equating the collapse time (determined by the eigenvalues) to the age of the Universe. For an isotropic collapse, they calculated explicitly $\mu_{\mathrm{th}} = 3.21$ (appendix A in Forero-Romero *et al.*, 2009). As gravitational collapse is often highly anisotropic, they used an empirical approach to determine the threshold and argued that $\mu_{\mathrm{th}} \approx 1$ can yield better web classifications than the original T-web, down to the megaparsec scale.

The T-web procedure and/or this extension have been used, for example, by Jasche *et al.* (2010b); Wang *et al.* (2012); Forero-Romero, Contreras & Padilla (2014); Nuza *et al.* (2014); Alonso, Eardley & Peacock (2015); Eardley *et al.* (2015); Forero-Romero & González (2015); Leclercq, Jasche & Wandelt (2015c); Zhao *et al.* (2015); Aung & Cohn (2016).

C.2.4.2 The V-web

Hoffman *et al.* (2012) reformulated the extended T-web procedure using the velocity shear tensor instead of the gravitational tidal tensor. More precisely, they use the eigenvalues $\mu_i^V(\mathbf{x})$ of the rescaled shear tensor defined by

$$\Sigma_{ij} \equiv -\frac{1}{2H(z)}\left(\frac{\partial \mathbf{v}_i}{\partial \mathbf{r}_j} + \frac{\partial \mathbf{v}_j}{\partial \mathbf{r}_i}\right). \tag{C.18}$$

This new scheme is generally referred to as the "V-web" and the rules are given in table C.3. Hoffman *et al.* (2012) showed that the two classifications coincide at large scales (where the gravitational and velocity fields are proportional) and that the velocity field resolves finer structure than the gravitational field at the smallest scales (sub-megaparsec). They empirically determined the threshold value $\mu_{\mathrm{th}}^V = 0.44$ to yield the best visualization of the geometrical characteristics of the four environments at $z = 0$.

Structure type	Rule
Void	$\mu_1^V, \mu_2^V, \mu_3^V < \mu_{\mathrm{th}}^V$
Sheet	$\mu_1^V, \mu_2^V < \mu_{\mathrm{th}}$ and $\mu_3^V > \mu_{\mathrm{th}}^V$
Filament	$\mu_1^V < \mu_{\mathrm{th}}^V$ and $\mu_2^V, \mu_3^V > \mu_{\mathrm{th}}^V$
Cluster	$\mu_1^V, \mu_2^V, \mu_3^V > \mu_{\mathrm{th}}^V$

Table C.3: Rules for classification of structure types according to the V-web procedure (Hoffman *et al.*, 2012).

The V-web has been used, for example, by Libeskind *et al.* (2013); Carlesi *et al.* (2014); Nuza *et al.* (2014); Lee, Rey & Kim (2014); Libeskind, Hoffman & Gottlöber (2014). In this thesis, we probe scales down to a few Mpc/h (the voxel size in our reconstructions or simulations). Therefore, we will be content with the original T-web procedure as formulated by Hahn *et al.* (2007a).

C.2.5 Implementation

This section gives details on how the T-web procedure is implemented when used in this thesis. First, the density contrast field is computed by assigning particles to the grid with a CiC scheme (see section B.3). It is transformed to Fourier space using a Fourier transform on the grid. At this point, if desired, the density field can be smoothed using a Gaussian kernel $K_{k_\mathrm{s}}(k) \equiv \exp\left(-\frac{1}{2}\frac{k^2}{k_\mathrm{s}^2}\right)$ (usually this step is bypassed in the projects described in this thesis). This corresponds to a mass scale M_s which is linked to the smoothing length $R_\mathrm{s} \equiv \frac{2\pi}{k_\mathrm{s}}$ by

$$R_\mathrm{s} = \frac{1}{\sqrt{2\pi}}\left(\frac{M_\mathrm{s}}{\bar{\rho}}\right)^{1/3}. \tag{C.19}$$

The reduced gravitational potential is estimated by solving the Poisson equation in Fourier space, $\tilde{\Phi}(\mathbf{k}) = G(\mathbf{k})\delta(\mathbf{k})$, where $G(\mathbf{k})$ is the Green function corresponding to the discretization adopted for the Laplacian. For the projects described in this thesis, we adopted the simple form $G(\mathbf{k}) = -1/k^2$ (with also a smoothing of short-range forces and two deconvolutions of the CiC kernel, see section B.4.1). Hence, the gravitational potential is given by the convolution

$$\tilde{\Phi}(\mathbf{x}) = (G * \delta)(\mathbf{x}), \tag{C.20}$$

or, if the density field had been smoothed, by

$$\tilde{\Phi}_{R_\mathrm{s}}(\mathbf{x}) = (G * K_{k_\mathrm{s}} * \delta)(\mathbf{x}). \tag{C.21}$$

We compute the components of the tidal tensor in Fourier space using $\mathcal{T}_{ab} = -\tilde{\Phi}(\mathbf{k})\mathbf{k}_a\mathbf{k}_b$, and transform them back to configuration space by inverse Fourier transform. In practice, only one Fourier transform is required to go from δ to $\mathcal{T}_{ab} \propto -\delta(\mathbf{k})\mathbf{k}_a\mathbf{k}_b/k^2$ (or $\mathcal{T}_{ab} \propto -\delta(\mathbf{k})\mathbf{k}_a\mathbf{k}_b K_{k_\mathrm{s}}(k)/k^2$). Finally, we compute the eigenvalues of the tidal tensor at each voxel of the grid and classify structures using the rules given in table C.1. In this fashion, every voxel of the density field gets assigned a flag corresponding to the structure type: T_0 for voids, T_1 for sheets, T_2 for filaments, T_3 for clusters.

The T-web classification takes a few seconds on 8 cores, for a typical density field used in this thesis ($L = 750$ Mpc/h, $N_\mathrm{v} = 256^3$).

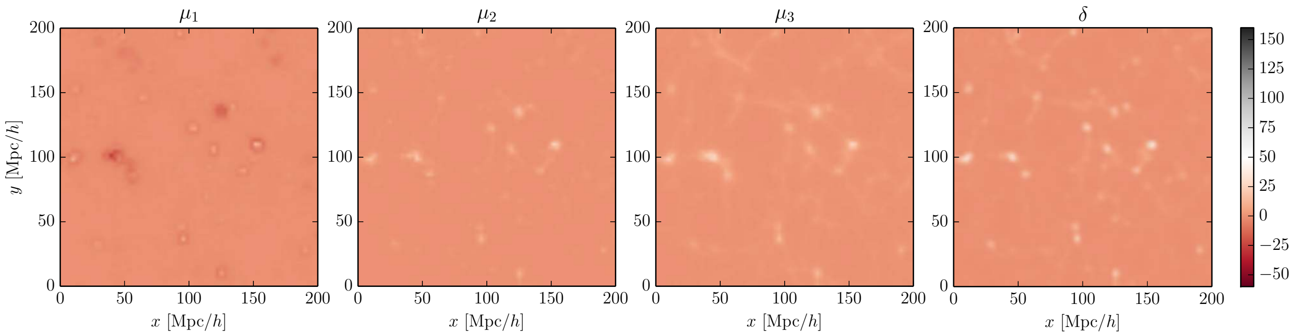

Figure C.1: Slices through the voxel-wise eigenvalues $\mu_1 \leq \mu_2 \leq \mu_3$ of the tidal field tensor in the final conditions of a dark matter simulation. The rightmost panel shows the corresponding slice through the final density contrast $\delta = \mu_1 + \mu_2 + \mu_3$ (equation (C.9)). See also figure 9.2 for comparison.

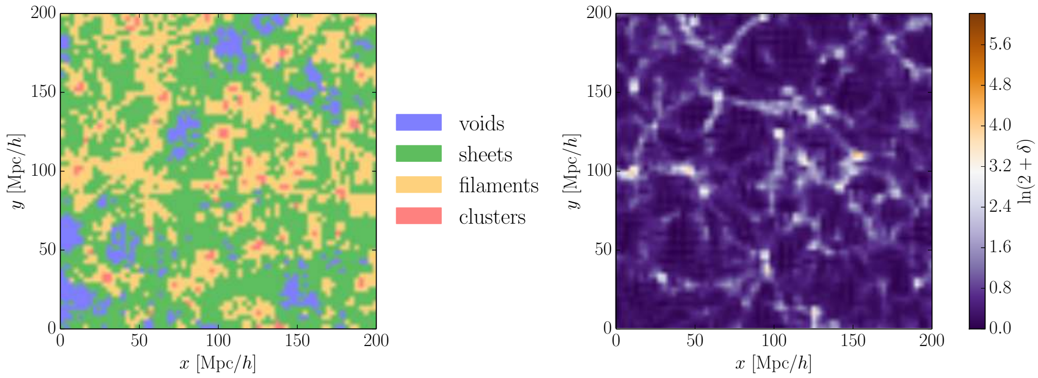

Figure C.2: *Left panel.* Classification of structures with the T-web procedure in the final conditions of a dark matter simulation. The color coding is blue for voids, green for sheets, yellow for filaments and red for clusters. *Right panel.* Dark matter density in the corresponding slice (for convenience, the quantity shown in $\ln(2 + \delta)$).

C.2.6 Example

As an example, in this section, we show the results of the T-web classification for a simulated density field.

The simulation contains 512^3 dark matter particles in a comoving box of 750 Mpc/h with periodic boundary conditions. The initial conditions have been generated at $z = 69$ using second-order Lagrangian perturbation theory. They obey Gaussian statistics with an Eisenstein & Hu (1998, 1999) power spectrum. The N-body simulation has been run to $z = 0$ with Gadget-2 (Springel, Yoshida & White, 2001; Springel, 2005). Particles are assigned to the grid using a CiC method. The cosmological parameters used are

$$\Omega_\Lambda = 0.728, \Omega_\mathrm{m} = 0.272, \Omega_\mathrm{b} = 0.045, \sigma_8 = 0.807, h = 0.702, n_\mathrm{s} = 0.961, \tag{C.22}$$

which gives a mass resolution of 2.37×10^{11} $\mathrm{M}_\odot/h$.

For clarity, we show slices through a 200 Mpc/h region of the simulation. Figure C.1 shows the eigenvalues of the tidal tensor and the density contrast. A slice through the corresponding voxel-wise classification of structures is shown in the left panel of figure C.2.