A New Multi-Scale Structure Finding Algorithm to Identify Cosmological Structure

Ali Snedden^{a,*}, Lara Arielle Phillips^a, Grant J. Mathews^a, Jared Coughlin^a,
In-Saeng Suh^a, Aparna Bhattacharya^a

^a*Center for Astrophysics
Department of Physics
225 Nieuwland Science Hall
University of Notre Dame
Notre Dame, IN 46556*

Abstract

We introduce a new structure finding algorithm that self-consistently parses large scale cosmological structure into clusters, filaments and voids. This structure finding algorithm probes the structure at multiple scales and classifies the appropriate regions with the most probable structure type and size. We show that it can identify the baryon fraction of intercluster medium and cosmological voids.

Keywords: Large Scale Structure, cosmology, structure finding, segmentation, clusters, filaments, voids

1. Introduction

Numerous large galaxy redshift surveys have revealed filaments and walls connecting clusters together with vast voids in between. This large scale structure (LSS) range in size from several to hundreds of megaparsecs. Location within the LSS, affects the gas density, temperature, and chemical enrichment history, the galaxy merger and harassment history and the orientation of galaxies [1, and references therein]. One of the primary motives in developing an algorithm that self-consistently identifies groups/clusters (henceforth jointly referred to as clusters), filaments and voids is to probe the effect of environment on galaxy evolution.

*Corresponding author. Tel.: 574-631-3583

URL: asnedden@nd.edu (Ali Snedden)

Preprint submitted to Journal of Computational Physics

May 3, 2015

The challenge is to find an optimum algorithm that can successfully identify clusters, filaments and voids in large observational and numerical data sets. Many different methods [2, 3, and references therein] have been proposed for identifying structures and their associated galaxies within large galaxy redshift surveys and cosmological simulations. Most algorithms follow one particular type of structure. Moreover, there are significant ambiguities in the definitions of what exactly constitutes a filament or void (clusters are well defined), both observationally and in simulations [4]. Hence, a comparison of structures found by different algorithms remains a challenging task [4].

Only a few algorithms [5, 6] follow multiple types of structure. We present a new structure finding algorithm adapted to solve this problem. It is similar to those presented by Refs. [5–7]. It utilizes the density curvature information provided by the eigenvalues of the Hessian matrix to find structures. However, unlike Refs. [5, 7], the structure finder presented here provides a self-consistent way for dealing with clusters, filaments *and* voids.

This new algorithm also has the advantage of being computationally more efficient than the structure finders of Refs. [5, 6]. We use the density obtained directly from the simulation output, without the need to compute either a Delaunay Tessellation Field Estimator [5] or the gravitational potential and particle trajectories [6].

2. Methods

The structure finding algorithm described here uses 2^{nd} order curvature information from the density distribution to parse the large scale structure. It is based on a vasculature segmentation algorithm [8] used in medical imaging. The algorithm uses the local curvature information to filter [8, 9] the filaments from other structures. To get this curvature information one can Taylor expand the density, $\rho(\mathbf{x}, \sigma)$, about the local position $\mathbf{x}_0$ at some characteristic scale, σ . The resulting expansion is

$$\rho(\mathbf{x}_0 + \delta\mathbf{x}, \sigma) \approx \rho(\mathbf{x}_0, \sigma) + \nabla\rho(\mathbf{x}_0, \sigma)_i \delta\mathbf{x}^i + \frac{1}{2}H(\mathbf{x}_0, \sigma)_{ij} \delta\mathbf{x}^i \delta\mathbf{x}^j + \dots \quad (1)$$

where $\nabla\rho(\mathbf{x}_0, \sigma)_i$ is the gradient of the density, and $H(\mathbf{x}_0, \sigma)_{ij}$ is the 3×3 Hessian matrix comprised of the second derivatives of the density. To calculate the spatial derivatives, the density field is convolved [9, 10] with a kernel composed of the derivatives of a Gaussian function along the desired dimension. Thus,

$$H(\mathbf{x}_0, \sigma)_{ij} = \int \nabla\rho(\mathbf{x}_0, \sigma)_i * \frac{\partial G(\mathbf{x} - \mathbf{x}_0, \sigma)}{\partial x_j} dx_j, \quad (2)$$

where

$$\nabla \rho(\mathbf{x}_0, \sigma)_i = \int \rho(\mathbf{x}, \sigma) * \frac{\partial G(\mathbf{x} - \mathbf{x}_0, \sigma)}{\partial x_i} dx_i \quad (3)$$

and

$$G(\mathbf{x} - \mathbf{x}_0, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\mathbf{x} - \mathbf{x}_0)^2}{2\sigma^2}}. \quad (4)$$

The characteristic Gaussian scale, σ , is iterated through a range of values to effectively probe structures of different sizes. The algorithm returns the maximum structure measure and the corresponding scale (σ) for each cell or voxel. Thus larger structures are probed best by wider Gaussian kernels and smaller structures by smaller Gaussian kernels.

The resulting eigenvectors yield three orthogonal axes with the eigenvalues giving the relative spatial curvature along each axis [9]. These eigenvalues are ordered by their relative magnitudes as

$$|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|. \quad (5)$$

The relative ratios of the eigenvalues determines the structural information. For example, if $|\lambda_1| \ll |\lambda_2|$ and $|\lambda_2| \approx |\lambda_3|$, then, one dimension has a relatively small curvature, while two dimensions have much higher curvature, which indicates a filament.

We expand this algorithm to find clusters and voids. The relative ratios of the eigenvalues are used to generate a unique structure measure for each type of structure. Following the logic from Refs. [8, 9], how cluster-like or filament-like a region of space is can be quantified by the structure measures

$$V_c = (1 - e^{-2|\lambda_1/\lambda_3|^2})(1 - e^{-2|F_{norm}|^2}) \quad (6)$$

and

$$V_f = (1 - e^{-2|\lambda_1/\lambda_2|^2})(1 - e^{-2|F_{norm}|^2}), \quad (7)$$

respectively, where F_{norm} is the Frobenius norm,

$$F_{norm} = (\lambda_1^2 + \lambda_2^2 + \lambda_3^2)^{\frac{1}{2}}. \quad (8)$$

The Frobenius norm term acts to wash out structure signal by suppressing small eigenvalue fluctuations. Building upon this logic, we introduced a new structure measure that is useful for voids

$$V_v = (1 - e^{-2|\lambda_1/\lambda_3|^2})(e^{-2|F_{norm}|^2}). \quad (9)$$

Here, the role of the Frobenius norm term has changed. Instead of decreasing the signal from small eigenvalue fluctuations, it enhances the signal. The reasoning is that regions of low density will be dominated by the small eigenvalue fluctuations not related to any physical structure. This provides a self-consistent method with which to investigate structures observed in the Universe.

In addition to equations 6 - 9, we also impose the criteria listed in Table 1 [9]. If the eigenvalues do not meet the criteria for the particular structure type, the appropriate structure measure is set to 0. This ensures that in the cases of clusters and filaments, we are indeed finding structures instead of the absence of structures (i.e. voids).

The structure measures are compared to each other to determine if a given region of space is more cluster-like, filament-like or void-like. In addition, we use a threshold for these measures, below which the structure type is categorized as unknown. For the purposes of this study, sheet-like structures are not explicitly traced because in the cosmological context they tend to be on much larger scales than those in the simulation presented here.

λ_1	λ_2	λ_3	structure type
−	−	−	cluster
≈ 0	−	−	filament
+	+	+	void

Table 1: The signs of the eigenvalues for the characterization of different structures. The clusters and filaments are negative due to the structure consisting of large density values in a sea of small values. In the void case, the structure being sought is the opposite of clusters, hence the opposite sign.

3. Results

In Section 3.1, the structure finding algorithm was run on a sub-volume of the dark-matter only Millennium Simulation [4, 11, 12]. This allows us to test the validity of our new void finding algorithm by benchmarking with respect to Ref. [4]. In Section 3.2, we explore the validity of the entire algorithm in finding clusters, filaments and voids in a cosmological simulation (including supernova feedback) and comparing it to observed results.

3.1. Millennium Simulation

A $(60 h^{-1}\text{Mpc})^3$ sub-volume containing 12,528,667 dark matter particles was extracted from the Millennium Simulation [11] for analysis. The associ-

ated 17,604 galaxies were also extracted from the semi-analytic galaxy catalogue generated by Ref. [12]. This was the same volume used by Aspen-Amsterdam Void Finder Comparison Project [4].

The structure finder was used to find voids within this region and then compared to the void finders described in Ref. [4]. This provides a benchmark from which to compare the validity of the void finding algorithm.

To convert the particle data to grid data, the dark matter particles were binned onto a 300^3 grid that was mapped to the sub-volume. The data was then smoothed using a Gaussian kernel with a standard deviation equal to $0.2 h^{-1}\text{Mpc}$. This slight smoothing was necessary to ensure the continuity of the derivatives.

Fig. 1 provides a 5 voxel thick view of a void overlaid on the dark matter and galaxy distribution. This compares favorably to Fig. 1 of Ref. [4], and illustrates that the central void has been correctly identified. We are also able to reproduce the same underdense void profile as shown in Fig. 3 of Ref. [4]. The void is significantly under-dense with respect to the overall volume. This verifies that the void finder works correctly.

3.2. Feedback Simulation

Having demonstrated that the structure finder can adequately find voids, we explore how well it can identify clusters, filaments and voids. A simulation was run using the smoothed particle hydrodynamics code, GADGET-2 [13], to which we have added the relevant physical processes: radiative cooling (CLOUDY version 10.0, last described in Ref. [14]), UV-background heating, star formation, supernovae feedback and metal enrichment which is fully described in Ref. [15].

To search for structures, the gas particle density from each snapshot was converted to a grid. The density at each grid intersection point was computed using the standard SPH kernel. The density for each voxel in the grid was calculated using the integration of the trilinear interpolating function constructed from the grid intersection points. This has the advantage of not artificially washing out features from non-physical smoothing.

There is no *a priori* knowledge of the best grid resolution to use when converting the particle data to a grid. The 256^3 initial gas particles from the simulation were converted to 128^3 , 256^3 and 512^3 grids. Each grid was filtered with respect to clusters, filaments and voids on scales $0.78\text{--}8.8h^{-1}\text{Mpc}$, $0.39\text{--}5.5h^{-1}\text{Mpc}$ and $1.9\text{--}19.5h^{-1}\text{Mpc}$ respectively. Fig. 2 gives a visual example of the filtering for a similar GADGET simulation. The clusters, filaments and voids are correctly identified. Approximately 7% of the gas mass is not assigned a structure type after the clusters, filaments, and voids are identified, in contrast with 37% of the volume. A deeper analysis of the unassigned component [16]

Figure 1: Void Outlined in Particle Data.

Illustration of a $1.0 h^{-1}\text{Mpc}$ thick (equal to 5 voxels) slice through a void region within a sub-volume of the Millennium Simulation. Red dots represent individual dark matter particles, blue points represent galaxies from the semi-analytic catalogue, the big black dot represents the geometric center of the void and the green filling represents the central void region

showed that the unassigned voxels were mostly found in the wings or outskirts of filaments. We also anticipate that there could be voids with scales smaller than the minimum probed here.

After filtering, the structure measures were compared to each other and the structure with the largest measure was selected. If the maximum structure measure was not above a threshold of 0.0038 (where the structure measure is normalized on a scale of 0 to 1), no structure type was assigned.

To determine which grid size to use for the 256^3 simulation, the cluster gas fraction from each of the different grid resolutions was compared to the observed gas fraction of the intracluster medium (ICM). In Ref. [17] the ICM mass fraction with respect to the baryons is reported to be $4 \pm 1.5 \%$. The 128^3 and 256^3 grid (see Table 2) agree with this to within the uncertainty. This indicates that an Eulerian grid corresponding to the simulation should have roughly the same number of cells as the SPH simulation has particles.

grid	cluster mass fraction
128^3	4.42%
256^3	3.38%
512^3	1.37%

Table 2: Cluster mass fraction with respect to the total baryon mass for the 128^3 , 256^3 and 512^3 grids.

(a) Unfiltered Gas Distribution.

(b) Cluster regions highlighted.

(c) Filament regions highlighted.

(d) Void regions highlighted.

Figure 2: GADGET-2 Simulation Filtering

A one voxel thick slice through a selected region of our modified GADGET-2 simulation. Fig. 2a shows the unfiltered gas density (scaled logarithmically.) Fig. 2b, Fig. 2c and Fig. 2d show the structure measure, normalized to 1 and scaled linearly, for clusters, filaments and voids respectively.

4. Discussion

A self-consistent structure finding algorithm, capable of simultaneously identifying clusters, filaments and voids, has been presented. The original algorithm was changed to include the additional void finding features. It has the advantage over other structure finders [2, 3, and references therein] in that it provides a self-consistent method for classifying structures. When compared to the only other structure finders [5, 6] in the literature capable of finding multiple types of structures, this algorithm provides simplicity in implementation.

The structure finder can successfully find the same void as the void finders tested in Ref. [4] and has a similar radial density distribution. This verifies that the void finder is behaving reasonably. When the structure finder was tested on a large, cosmological simulation it reproduced the correct fractional amount of the ICM. Future studies will completely follow the gas phase evolution and the star formation history of the structures [16].

5. Acknowledgments

The subvolume data set of the Millennium simulation was generously provided by Frazer Pearce. The original filament finding algorithm was generously provided by Maxime Descouteaux. This research was supported in part by the Notre Dame Center for Research Computing. Work also supported in part by DOE grant DE-FG02-95-ER40934.

6. References

- [1] Y. Zhang, X. Yang, A. Faltenbacher, V. Springel, W. Lin, H. Wang, The Spin and Orientation of Dark Matter Halos Within Cosmic Filaments, *ApJ*.706 (2009) 747–761. arXiv:0906.1654, doi:10.1088/0004-637X/706/1/747.
- [2] E. Platen, R. van de Weygaert, B. J. T. Jones, G. Vegter, M. A. A. Calvo, Structural analysis of the SDSS Cosmic Web - I. Non-linear density field reconstructions, *Mon. Not. R. Astron. Soc.*416 (2011) 2494–2526. arXiv:1107.1488, doi:10.1111/j.1365-2966.2011.18905.x.
- [3] E. Tempel, R. S. Stoica, V. J. Martínez, L. J. Liivamägi, G. Castellan, E. Saar, Detecting filamentary pattern in the cosmic web: a catalogue of filaments for the SDSS, *Mon. Not. R. Astron. Soc.*438 (2014) 3465–3482. arXiv:1308.2533, doi:10.1093/mnras/stt2454.
- [4] J. M. Colberg, F. Pearce, C. Foster, E. Platen, R. Brunino, M. Neyrinck, S. Basilakos, A. Fairall, H. Feldman, S. Gottlöber, O. Hahn, F. Hoyle, V. Müller, L. Nelson, M. Plionis, C. Porciani, S. Shandarin, M. S. Vogeley, R. van de Weygaert, The Aspen-Amsterdam void finder comparison project, *Mon. Not. R. Astron. Soc.*387 (2008) 933–944. arXiv:0803.0918, doi:10.1111/j.1365-2966.2008.13307.x.
- [5] M. A. Aragón-Calvo, B. J. T. Jones, R. van de Weygaert, J. M. van der Hulst, The multiscale morphology filter: identifying and extracting spatial patterns in the galaxy distribution,

- Astronomy and Astrophysics 474 (2007) 315–338. arXiv:0705.2072, doi:10.1051/0004-6361:20077880.
- [6] O. Hahn, C. M. Carollo, C. Porciani, A. Dekel, The evolution of dark matter halo properties in clusters, filaments, sheets and voids, *Mon. Not. R. Astron. Soc.* 381 (2007) 41–51. arXiv:0704.2595, doi:10.1111/j.1365-2966.2007.12249.x.
 - [7] N. A. Bond, *Crawling the cosmic network: An exploration of filamentary structure*, Ph.D. thesis, School: Princeton University (2008).
 - [8] M. Descoteaux, L. Collins, K. Siddiqi, A Multi-Scale Geometric Flow for Segmenting Vasculature in MRI, in: *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, Vol. 3117 of *Lecture Notes in Computer Science*, 2004, pp. 169–180.
 - [9] A. F. Frangi, W. J. Niessen, K. L. Vincken, M. A. Viergever, Multiscale Vessel Enhancement Filtering, in: *Medical Image Computing and Computer-Assisted Intervention - MICCAI'98*, Vol. 1496 of *Lecture Notes in Computer Science*, 1998, pp. 130–137.
 - [10] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, M. A. Viergever, Scale and the differential structure of images, *Image and Vision Computing* 10 (1992) 376–388.
 - [11] V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, F. Pearce, Simulations of the formation, evolution and clustering of galaxies and quasars, *Nature* 435 (2005) 629–636. arXiv:arXiv:astro-ph/0504097, doi:10.1038/nature03597.
 - [12] D. J. Croton, V. Springel, S. D. M. White, G. De Lucia, C. S. Frenk, L. Gao, A. Jenkins, G. Kauffmann, J. F. Navarro, N. Yoshida, The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies, *Mon. Not. R. Astron. Soc.* 365 (2006) 11–28. arXiv:arXiv:astro-ph/0508046, doi:10.1111/j.1365-2966.2005.09675.x.
 - [13] V. Springel, The cosmological simulation code GADGET-2, *Mon. Not. R. Astron. Soc.* 364 (2005) 1105–1134. arXiv:astro-ph/0505010, doi:10.1111/j.1365-2966.2005.09655.x.
 - [14] G. J. Ferland, K. T. Korista, D. A. Verner, J. W. Ferguson, J. B. Kingdon, E. M. Verner, *CLOUDY 90: Numerical Simulation of Plasmas and Their Spectra*, *Publications of the Astronomical Society of the Pacific* 110 (1998) 761–778. doi:10.1086/316190.
 - [15] A. Snedden, *Evolution of Baryons in Cosmic Large Scale Structure*, Ph.D. thesis, Ph. D. thesis, University of Notre Dame (2015) (Jan. 2015).
 - [16] A. Snedden, J. Coughlin, L. A. Phillips, G. Mathews, I.-S. Suh, *Star Formation and Gas Phase History of the Cosmic Web*, submitted to *Mon. Not. R. Astron. Soc.* arXiv:1412.7050.
 - [17] J. M. Shull, B. D. Smith, C. W. Danforth, The Baryon Census in a Multiphase Intergalactic Medium: 30% of the Baryons May Still be Missing, *ApJ* 759 (2012) 23. arXiv:1112.2706, doi:10.1088/0004-637X/759/1/23.