The Segmented Universe:
Identifying Cosmic Voids with a Multi-Scale Geometric Flow

Andrew Miller

2010 NSF/REU Program
Physics Department, University of Notre Dame

Advisor: Dr. Lara Arielle Phillips
Abstract

The complex, filamentary nature of large-scale dark matter and density structure in the universe is a prominent feature of both redshift surveys and large N-body simulations of cosmic evolution. Here, we present a quantitative method for the analysis of such structure though the application of a medical imaging algorithm to dark matter and semi-analytical model galaxy data from the Virgo Consortium's 2005 Millennium Simulation (MS). The algorithm, a multi-scale geometric flow for segmenting vasculature in proton density images of the human brain, originally identified regions of vessel-like structure in an intensity field using level set methods. Due to the striking similarity between the types of structure observed in the universe on cosmic scales and the type of organization exhibited by the vasculature of the brain, we theorized that this method of structure analysis would prove an accurate means of classification for formations of cosmic dark matter overdensity and the accompanying density voids. We have modified the algorithm to identify sheet-like, clump-like, and void-like structure in addition to filament-like features in a density field. We focus on regions of cosmic void to determine the robustness of this segmentation method from a systematic comparison of its results with those of previously published void-finding algorithms. To this end, we extract the dark matter halo distribution from a subvolume of the MS and convert it to a three-dimensional image of variable cell resolution. We distribute the mass of each halo according to the Navarro, Frenk, & White (NFW) mass-density profile and write the data set to a Medical Image NetCDF (MINC) volume. We apply the segmentation algorithm to this MINC volume and analyze the initial results.
Introduction

The existence of structure in our universe has been known almost since the discovery of galaxies remote from our own Milky Way.\textsuperscript{1} Abell presented the first evidence of large associations of galaxy clusters, providing an effective means for a systematic phenomenological investigation of the large-scale distribution of matter in the universe.\textsuperscript{2} Since then, large redshift surveys, such as the Sloan Digital Sky Survey (SDSS) have revealed in great detail the nature of this large-scale distribution of galaxies: the clusters connected by filaments and sheets, and the great regions of void in between.

Voids in cosmic structure are generally defined as underdense regions in the current mass distribution of the universe, often with very steep edges.\textsuperscript{3} Readily apparent as the counterpart to the overdense filaments observed both in redshift surveys and in theoretical cold dark matter (CDM) numerical simulations\textsuperscript{4,5,6,7}, voids have become a prominent means of large-scale structure (LSS) analysis. The current CDM cosmological model (augmented with the cosmological constant $\Lambda$, representing the dark energy density which governs the rate of expansion of the universe) holds that such observed cosmic structure has its origins in weak density fluctuations in the early ages of the otherwise homogenous universe (observable today in the cosmic microwave background radiation) which were exacerbated through gravitational clustering.\textsuperscript{8,1,7} Voids themselves grow around regions of underdensity as a result of the gravitational collapse of dark matter particles in the initially nearly homogenous CDM distribution of the early universe into filaments and clusters around regions of original weak overdensity. These overdense areas develop into interstitial nodes\textsuperscript{9} between a packing of roughly spherical voids, towards which matter continues to condensate, causing the fluctuations of large-scale mass distribution of the universe to become more pronounced with time. Within
regions both of void and of populated space, matter density is not homogenous, being collected into local regions of underdensity and overdensity analogous to the larger-scale structure.

Although it is possible to model the initial growth of density perturbations analytically, the nonlinear nature of gravitational collapse and hierarchical structure development necessitates direct numerical simulation. Due to the dominance of CDM, which subject only to the forces of gravitational attraction, over regular matter the mass distribution of the universe may be modeled as a hydrodynamic-free set of discrete point particles. We use the Millennium Simulation (MS) and Millennium-II Simulation, which were carried out by the Virgo Consortium in the years 2005 and 2009 respectively, and follow approximately $1.0078 \times 10^{10}$ simulation particles from redshift $z = 127$ to the present. This most recent family of high-resolution N-body $\Lambda$CDM simulation represents the largest pure dark matter efforts to date.

Attempts have been made to systematically quantify the LSS of the universe seen in redshift surveys and numerical simulation, notably Regős and Geller, Babul and Starkman, and Mecke, Buchert, and Wagner. A wide variety of void-finding algorithms have sprung from these early efforts, motivated by the importance of cosmic voids to a full understanding of cosmic structure, a greater knowledge of cosmological information and galaxy formation. The *Aspen-Amsterdam Void-finder Comparison Project*, described as the first such systematic study in this area, evaluates thirteen void-finding algorithms from different groups on a small subvolume of the MS, and gives a qualitative and quantitative analysis of the differences of each.

We present a novel void-finding algorithm which is an adaptation of a computer vision algorithm designed to detect blood vessels in proton density images of the human brain and apply it to the MS dark matter distribution.
Simulation and Extraction Procedure

We obtain the simulation outputs from the MS, which comprises a $500h^{-1} Mpc$ per side cubic region with periodic boundary conditions, where the total matter density is given $\Omega_m = 0.25$, the baryonic matter density $\Omega_b = 0.045$, the dark energy density $\Omega_\Lambda = 0.75$, the Hubble constant $h = 0.73$, and the normalization of the power spectrum $\sigma_8 = 0.9$. Two separate semi-analytic galaxy formation models were implemented in the MS to populate the CDM substructure.\textsuperscript{14,15}

For this work, we use the CDM halo distribution and the L-Galaxies $z = 0$ semi-analytic galaxy catalogue\textsuperscript{14}. From the SUBFIND\textsuperscript{16,17} halo table we extract the spatial coordinates and the mass within the radius where the halo has an overdensity 200 times the critical density of the simulation of all halos at $z = 0$ in a $60h^{-1} Mpc$ per side subvolume of the MS. We extract a total of 4277 halos with $M_{200} > 10^{11} h^{-1} Mpc$, comprising 7655905 simulation particles, giving our subsample an underdensity $\delta = \rho / \bar{\rho} - 1 = -0.56$ (where $\rho$ is the density of our MS subvolume and $\bar{\rho}$ is the density of the entire MS). In this subvolume we extract the spatial coordinates, stellar and cold gas mass, and BVRIK dust-corrected magnitudes of the 39753 galaxies with $B < -10$ present at $z = 0$ in the semi-analytical L-Galaxies catalogue.

NFW Profiling

Since the MS gives as the position of each halo only the coordinates of the constituent particle with the least potential energy, we distribute the mass of each halo into a volume to reconstruct the spatial extend of the halo. We use the Navarro, Frenk, & White\textsuperscript{18} (NFW) mass-density profile, a simple formula with two free parameters:

$$\rho(r)/\rho_{crit} = \delta_c \left[ \left( r/r_s \right)(1 + r/r_s)^2 \right].$$

(1)
The critical density to collapse is given in the MS: \( \rho_{\text{crit}} = \frac{3H_o^2}{8\pi G} \), where \( G \) is the gravitational constant, \( H_o \) is Hubble’s constant, and \( r \) is the radial distance from the center of the halo’s mass. \( r_s \) is a scaled, characteristic radius, \( r_s = r_{200}/c \), defined in terms of the dimensionless halo concentration parameter, \( c \), and \( r_{200} \), the radius inside which the halo has an overdensity 200 times \( \rho_{\text{crit}} \), such that the mass inside that radius \( M_{200} = 200\rho_{\text{crit}}(4\pi/3)r_{200}^3 \). \( \delta_c \) is a “characteristic density contrast”, a dimensionless parameter:

\[
\delta_c = \frac{(200/3)c^3}{\left[\ln(1+c) - c/(1+c)\right]} . \tag{2}
\]

The halo concentration parameter \( c \) is specific to each halo. It has a well-defined, though weak, correlation to mass, generally decreasing with increasing mass.\(^{19}\) We use the power law derived by Neto\(^{19}\) from a sample of MS halos:

\[
c_{200} = 4.67 \left( \frac{M_{200}}{10^{14} h^{-1} M_{\odot}} \right)^{-0.11}, \tag{3}
\]

where \( M_{\odot} \) is the mass of the Sun. Using eqs. (1) – (6), a radial density profile only dependent upon \( M_{200} \) is found, and by integrating over a spherical volume we obtain the spatial dependence of the of the halo mass function

\[
M(r) = 4\pi \int_0^r \rho(r) r^2 dr = 4\pi \rho_{\text{crit}} \delta_c r_s^3 \left[ \frac{r_s}{r + r_s} + \ln(r + r) \right], \tag{4}
\]

whereby the mass of the halo may be distributed over a number of voxels in the MINC volume. Figure 1 displays the radial density profile, mass distribution, and volume distribution of a moderately-sized halo of \( M_{200} = 200 \times 10^{10} h^{-1} M_{\odot} \). Figure 2 shows the halo, at an arbitrary position in an array with cell size \( (0.05 h^{-1} Mpc)^3 \), distributed over a volume.
**Implementation**

The MS data are obtained in comma-separated variable (CSV) format, and a C program was written to convert the data into a three-dimensional Medical Image NetCDF (MINC) volume of variable cell resolution. The NFW profile is applied to each halo. The segmentation program is applied to the MINC volume in three separate stages filtering for filaments, sheets, and clumps,
respectively, each producing a new MINC volume holding the corresponding structure measure for each cell. The “ncdump” shell command in the NetCDF library allows extraction of MINC data in common data language format (CDL). A C program was written to read CDL data and extract it for analysis. We use existing tools to visualize the resulting volumes. Figure 3 displays a $15h^{-1}Mpc$ slice of a MINC data volume of $480^3$ cells after the NFW profile has been applied.

![Figure 3](image)

*Figure 3*

A slice of thickness $15h^{-1}Mpc$ in the $z$ dimension through the center of the subvolume extracted from the MS. The image shows all galaxies in slice plotted in green (light) over dark matter, distributed according to NFW profile, plotted in red (dark).according to the NFW profile. Axes give voxel coordinate in the MINC volume. Vertical lines through plot are artifacts of graphing program.

**Segmentation Algorithm**

The multi-scale geometric flow, employed here as the basis for a void-finding algorithm, was developed by Descoteaux, Collins, and Siddiqi at the McConnell Brain Imaging Centre of the Montreal Neurological Institute. The algorithm uses the eigenvalues of the Hessian matrix with
Frangi’s vesselness measure to identify tubular structure in a density or intensity field. Descoteaux’s algorithm was modified to also identify sheet-like and clump-like structures in addition to filaments in baryonic density data.

The full, modified algorithm may be described as three processes: structure filtering, measure propagation, and structure bounding. The structure filtering process, which is our focus here, first convolves the density distribution in the MINC volume with a Gaussian, \( G(\sigma) \), at multiple scales. Then, using the first and second order information in the Taylor expansion of the vector field around each point

\[
\rho(x_0 + \delta x_0, \sigma) = \rho(x_0, \sigma) + \nabla \rho \cdot \delta x_0 + H_{ij} \delta x_0^i \delta x_0^j + \ldots
\]  

(5)

and curvature is measured at each scale level. \( H \) is the Hessian matrix, defined for a function \( f(x_1, x_2, x_3) \) in three dimensions as

\[
H = \begin{bmatrix}
\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\
\frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\
\frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3^2}
\end{bmatrix}.
\]  

(6)

The eigenvalues of \( H \) provide information about the level and type of local structure at each point. Setting the eigenvalues \( |\lambda_1| \leq |\lambda_2| \leq |\lambda_3| \), four possible conditions arise: (1) Filament-like structure if

\[
|\lambda_1| = 0, |\lambda_2| = |\lambda_3| >> 0,
\]

(2) Sheet-like structure: \( |\lambda_1| = |\lambda_2| = 0, |\lambda_3| >> 0 \), (3) Clump-like structure:

\[
|\lambda_1| = |\lambda_2| = |\lambda_3| >> 0,
\]

and (4) Noise-like structure: \( |\lambda_1| = |\lambda_2| = |\lambda_3| = 0 \). Four corresponding parameters may thus be defined

\[
R_A = |\lambda_2/\lambda_3|,
\]

(7)

\[
R_C = |\lambda_1/\lambda_2|,
\]

(8)
\[ R_D = |\lambda_1/\lambda_3|, \quad (9) \]
\[ S = \left(\lambda_2^2 + \lambda_3^2 + \lambda_1^2\right)^{1/2}/S_{\text{max}}. \quad (10) \]

A low \( R_A \) indicates the presence of sheet-like structure, a low \( R_C \), the presence of filament-like structure, and a high \( R_D \), the presence of clump-like structure. The fourth parameter, the Frobenius norm, is used to eliminate random noise effects.\(^{20}\) Three measures of local structure are calculated for each point from the above values:

Sheetness measure: \( M_1(\sigma) = \left(e^{-\alpha_1 R_A^2} - e^{-\alpha_1}\right)\left(1 - e^{-\beta S^2}\right), \quad (11) \)

Filament measure: \( M_2(\sigma) = \left(e^{-\alpha_2 R_C^2} - e^{-\alpha_2}\right)\left(1 - e^{-\beta S^2}\right), \quad (12) \)

Clumpiness measure: \( M_3(\sigma) = \left(e^{-\alpha_3 R_C^2}\right)\left(1 - e^{-\beta S^2}\right). \quad (13) \)

We adapt the algorithm to identify void-like structures in the CDM distribution and semi-analytic galaxy catalogue of the MS through three separate methods. In the first method, we identify voids as absences of positive structure. Those cells in the MINC data volume assigned by the filtering process with high sheetness, filament, or clumpiness measures, are identified, and all other cells are designated as being void-like. In the second method, we identify voids as regions of positive clump-like or group-like shape. The density distribution of the data volume is first inverted, and then the segmentation filter run as normal. The cells assigned by the filter with high clumpiness measures are designated as belonging to a void. In the third method, yet untried, we will create a new voidness measure, based on the Eigen values of the Hessian in a manner analogous to the existing sheetness, filament, and clumpiness measures.
Conclusions

We have successfully applied a novel multi-scale geometric flow void-finding algorithm to the CDM halo distribution of the MS. We have completed a full pipeline to implement the segmentation program on MS data. In addition, a set of supplementary plotting and sorting programs have been written to provide information about and visualization of data at the various stages of the process. Analysis of the MS data using the filament, sheetness, and clumpiness structure measures to identify void structure, must follow. The three methods of adapting the segmentation program should be thoroughly tested and the robustness of our void-finding methods be measured against contemporary void-finding efforts. The implementation process has been designed with such testing in mind. To evaluate our algorithm by the current standard, we seek to apply the segmentation algorithm to the same MS subvolume and accord it the same analysis as that of the Aspen-Amsterdam Void-finding Project. In the future, the segmentation algorithm may be extended to include the final two stages of structure identification: measure propagation, and structure bounding. Although not discussed here, these two processes have the potential to improve the ability and accuracy of the void-finding capabilities of the segmentation program.

References