Exploring the geometry, topology and morphology of large scale structure using Minkowski functionals

Jatush V. Sheth[1] and Varun Sahni[2,*]

[1]Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411 007, India and Max Planck Institute für Astrophysik, D-85748, Garching, Germany
[2]Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411 007, India

Modern redshift surveys such as the 2 degree field Galaxy Redshift Survey (2dFGRS) and the Sloan Digital Sky Survey (SDSS) reveal fully 3-dimensional distribution of 10^{5-6} galaxies over a large cosmological volume $\simeq 0.1-1$ $[h^{-1}\ \mathrm{Gpc}]^3$. It is well established that the galaxies in these surveys show strong clustering. To the eye, the galaxies are distributed along sheet-like and/or filamentary superclusters. The *CfA Great Wall*, *Southern Great Wall* and the recently discovered *SDSS Great Wall* are the most popular superclusters of this kind. The superclusters are interwoven with one another, leaving $\geq 70\%$ of volume devoid of any visible matter. This volume is occupied by *voids*.

One is motivated to test the theoretical predictions for the clustering of galaxies against rich datasets resulting from these redshift surveys. To this end, several workers have recently proposed and developed an approach to quantify the large scale structure (LSS) by studying the geometry and topology of the superclusters and voids. Concretely, this can be achieved by evaluating the Minkowski functionals (MFs) for LSS-datasets. The MF-based approach further provides an unbiased description of the shapes and sizes of the elements of LSS, i.e. the superclusters and voids. This eventually leads to a framework within which to quantify LSS and compare outputs from simulations with redshift surveys.

In this review we give a summary of the progress made in this direction. After reviewing the status of observations and of numerical simulations, we comment upon the nature of bias which itself serves as a link between theoretical predictions and observations. Next we introduce MFs and give sufficient motivation for employing them in cosmology. We summarize the methods developed for efficient numerical estimation of MFs for cosmological datasets. Next we list out several important results obtained using these methods. Specifically, we stress the discriminatory power of MFs and of the derived morphological statistics, the Shapefinders. Shapefinders are specifically important to study the shapes and sizes of the superclusters and voids. Several successes of Shapefinders are highlighted here. We further note some of the important effects of scale-dependent bias which are brought out by an MF-based study of the mock catalogues of galaxies. Such effects, we note, should be accounted for before comparing theoretical models with observations.

I. Introduction

During the past decade, CMB experiments such as WMAP[1], type-Ia supernova searches and observations of galaxy clustering, have helped in turning cosmology into a precision science. The combined set of measurements point to a *flat*, low density ($\Omega_m \simeq 0.3$), accelerating universe dominated by dark energy ($\Omega_{\mathrm{DE}} \simeq 0.7$). Modern redshift surveys aim to complement the above information about background cosmological parameters by providing a detailed map of the distribution of baryonic matter on large scales. Redshift surveys have so far been primarily used for two purposes: (1) to get independent constraints on the matter density and several other cosmological parameters (see below) and (2) to measure the abundance and clustering properties of a variety of galaxies. The former exercise provides important consistency tests for cosmic parameters estimated using other methods, whereas the latter gives precious feedback to the physics of galaxy formation. In addition, it is also important to test whether the observed large-scale clustering properties of galaxies are consistent with the predictions of the preferred cosmological model(s). In order to carry out such a test, one needs to quantify the information content hidden in redshift space distribution of galaxies. In this article we review the methods developed to study the geometry, connectedness and morphology of large scale structure (LSS) in redshift surveys, which can eventually help us meet such a goal. Such an approach gives valuable insights into the large-scale distribution of matter, and provides a framework within which to test the paradigms of gravitational instability and of the Gaussianity of the primordial density field.

Impressive redshift surveys have been undertaken recently to map out the 3-dimensional distribution of galaxies. In the next section, we shall outline the history of redshift surveys,

*This was the last paper written by Jatush Sheth, who met his tragic death in a road accident on 27 November 2004.

*For correspondence. (e-mail: varun@iucaa.ernet.in)

and focus on two major recent surveys, 2dFGRS and SDSS. LSS in redshift surveys exhibits phenomenally rich and complex texture, wherein sheet-like and/or filamentary superclusters intersperse with large, empty regions called *voids*. Several salient aspects of the LSS of the Universe will be dwelt upon and we shall stress why it is important to study the morphology of LSS[2]. In subsequent sections we shall show that the complex morphological features of LSS are objectively quantifiable and that morphological information supplements traditional approach of understanding LSS using its n-point correlation functions and can be employed to confront theories with observations in an integrated manner. Statistical methods which probe the geometry and topology of LSS include percolation analysis[3,4], genus curve[5,6], minimal spanning trees[7], void probability function[8], Voronoi tesselations[9], as well as more recently introduced Minkowski functionals (MFs)[10,11]. For an excellent review of these different approaches, the reader may refer to refs 12–15. In this article, our focus will be on MFs.

II. Redshift surveys

Ever since sky maps of galaxy counts, e.g. Lick Catalogue[16] revealed a rich pattern in LSS of galaxies in projection, it was realized that complementing this information with the redshifts of individual galaxies was essential to get a fully 3-dimensional view of the Universe. In such an exercise, one turns the redshift of a galaxy into its distance with the help of Hubble's law[17], and is thus able to study the *redshift space* distribution of galaxies.

Various small surveys were carried out in the late 1980s and mid-1990s in a controlled fashion; their positive scientific outcome providing impetus for the initiation of bigger ones. Of these, perhaps the most influential was the Center for Astrophysics Survey[18]. This survey had an angular coverage of $6° \times 120°$ and a depth of $\simeq 150\ h^{-1}$ Mpc. The survey measured redshifts of about 2400 galaxies. The famous Great Wall centered around the Coma cluster was the most striking feature of this survey. The slice showed very clearly the 'bubbly' nature of LSS with voids having sharp boundaries; the largest void having a diameter of $\simeq 50\ h^{-1}$ Mpc. Subsequent surveys including follow-up CfA slices and the ESO Southern Sky Survey[19] amply confirmed the impression given by the first CfA slice. Indeed, the Southern Sky Redshift Survey (SSRS) discovered a second Great Wall in the southern sky. The emerging picture of filamentary and sheet-like superclusters[20] encircling frothy voids prompted cosmologists to seek an explanation from theory, and this interplay between theory and observations is only now maturing into a truly quantifiable discipline. Another important survey, the Optical Redshift Survey (ORS)[21] had a depth of $80\ h^{-1}$ Mpc, but attempted a complete sky coverage (except for the *zone of avoidance*). This survey measured 8500 redshifts in total, and was heavily used to estimate luminosity functions, galaxy correlations, velocity dispersions, etc. In the mid-1990s, the Stromlo-APM redshift survey[22] and the Durham/UKST redshift survey[23] were focussed towards the southern sky and led to fruitful results on correlation functions in real and redshift space, power spectra, redshift distortions, cosmological parameters, bias, etc. However, the galaxy redshifts in all these surveys never exceeded ~10000.

The Las Campanas Redshift Survey (LCRS)[24] consisted of six $1.5° \times 80°$ slices (3 each in the northern and southern galactic hemisphere), went to the depth of $\simeq 750\ h^{-1}$ Mpc ($z \sim 0.25$) and recorded redshifts of about 25,000 galaxies. LCRS presented an order-of-magnitude improvement over earlier surveys. This was the first deep survey of sufficient volume which could test ideas such as the statistical homogeneity of the Universe on large scales. Other important results from analysing LCRS included the luminosity function, second and third-order correlation functions, power spectra, etc. LCRS is perhaps the *only* large survey wherein both the topology and morphology of LSS have been extensively investigated[25,27] and used to confront observations with theoretical models[28].

The 2 degree Field Galaxy Redshift Survey (2dFGRS)

The 2dF multi-fiber spectrograph on the 3.9 m Anglo-Australian Telescope is a very impressive facility which can obtain spectra for up to 400 objects simultaneously over a field of view with 2° diameter. The spectrograph was put to use to sample two contiguous constant-declination strips of the sky, one in the northern galactic hemisphere [with angular coverage of $10° \times 90°$ in (δ, α)] and the other in the southern galactic hemisphere [with angular coverage $\sim 15° \times 90°$ in (δ, α)]. Working in the B_J band, the survey reached a depth of $z = 0.3$ with median redshift of 0.11, and has compiled redshifts of about 2.5×10^5 galaxies which are brighter than an extinction corrected magnitude[29] of $b_J = 19.45$. The survey is already complete and the resulting galaxy correlation functions, redshift distortions and pairwise velocity dispersions[30] demonstrate the superb quality of the data set. The science output of this survey and its overall cosmological relevance have been excellently reviewed in ref. 31. We shall discuss here some of the results which are most relevant to our present discussion.

One of the main science goals of 2dFGRS was to get improved estimates of the 3-dimensional power spectrum and the two-point galaxy–galaxy correlation function[32–35]. Since 2dFGRS sampled a cosmological volume as large as $\simeq 0.1[h^{-1}\ \text{Gpc}]^3$, it became possible to measure these quantities over a large range of scales $0.01 \leq k \leq 0.4\ h\ \text{Mpc}^{-1}$ with an unprecedented accuracy[36,37]. Assuming a constant bias between dark matter and baryons on scales $k \leq 0.15\ h\ \text{Mpc}^{-1}$, Percival *et al.*[36] compared the shape of the power spectrum with the primordial power spectrum predicted by linear perturbation theory of structure forma-

tion[36]. This allowed them to constrain the values of Ω_m, Ω_B after priors on the power spectrum index n and the Hubble constant h had been applied ($\Omega_m = 0.26 \pm 0.03$, $\Omega_B = 0.044 \pm 0.016$). Apart from this, 2dFGRS was employed to obtain estimates for the luminosity function of early and late type galaxies[38], and to study the luminosity dependent clustering of galaxies as measured using the correlation function statistic[39]. The latter study provided clues about the large scale bias between baryonic and dark matter and revealed that the former might follow the latter through a linear relationship (also see ref. 40).

Cosmological parameter estimation is clearly one of the key results gleaned from large redshift surveys such as 2dFGRS. However, in our view several assumptions made in the process remain to be adequately justified, for instance, the assumption of linear, constant bias used in the exercise mentioned above is clearly inadequate. In fact, the study of biasing of galaxies with respect to the underlying mass distribution has its own history, and the issue has not yet settled completely. Below we shall briefly discuss the currently established ideas about bias.

Figure 1 shows the distribution of galaxies in the flux-limited 2dFGRS slices. To the eye, galaxies are distributed anisotropically, with sharp qualitative similarities with the distribution of dark matter in N-body simulations (see below). The 3-dimensional power-spectrum, in a more appropriate form $\Delta^2(k) = dP(k)/d\ln(k)$, measures the amount of power per unit interval in the logarithmic bin around wave number k. In simple terms, this is a statistical measure of the density fluctuations within spheres of radius $R \sim 1/k$. Clearly, in a measure such as this, the information on length-scales $r < R$ is *averaged out*. Hence, one learns relatively little about the nature of the *actual* distribution of galaxies. Describing (and explaining) the visually rich filamentary pattern of superclusters of galaxies and quantifying the complex nature of the LSS is another, *equally* important application of redshift surveys. This problem has not been adequately addressed despite considerable recent progress[41–47]. Clearly, the contribution of higher order correlation functions is an important ingredient in such a program. Since higher order correlations are difficult to measure in practice, it would be useful if we could quantify the connectedness and non-Gaussianity of LSS as well as the shapes and sizes of individual superclusters and voids using geometrical diagnostics. The methods reviewed in the present article are intended to fill this gap in our knowledge and to complement more traditional, correlation function based approaches in developing a quantifiable picture of the LSS and its properties.

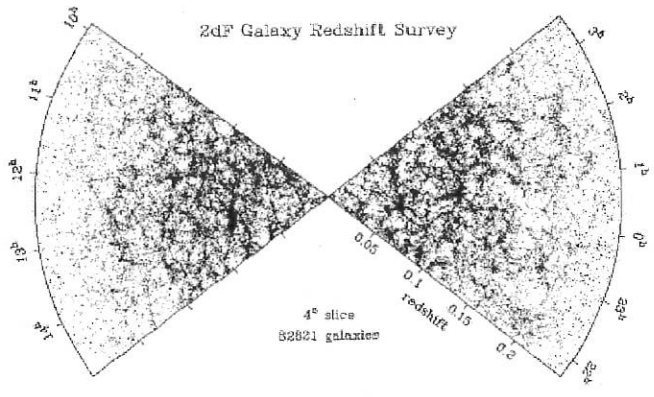

Figure 1. Flux-limited 2dFGRS slices in northern and southern galactic hemisphere are shown. The superclustervoid network visually stands out, and can be properly understood by computing either a hierarchy of correlation functions or by alternate, geometric and topological quantifiers. In this article, we review methods to delineate individual superclusters and voids and describe their properties by using topological and geometrical diagnostics such as Minkowski functionals and Shapefinders.

Sloan digital sky survey

The Sloan Digital Sky Survey (SDSS) is derived from a dedicated 2.5 m telescope. The initial photometric program has been to measure the positions and brightness of about 10^8 objects in π steradian (almost 1/4th) of the sky. The survey is centered on the northern galactic pole and has an elliptical angular coverage of $130° \times 110°$, where the semi-major axis runs along a line of constant right-ascension.

Follow-up spectroscopy is planned to give redshifts of about 10^6 galaxies and 10^5 quasars. At the time of writing, the SDSS team has made its 2nd data release, available for free download[48]. It contains redshifts of $\simeq 3° \times 10^5$ galaxies, with the farthest galaxy at a redshift of ~0.3. (See http://cas.sdss.org/astro/en/tools/search/SQS.asp.)

The limiting magnitude of the SDSS survey is in the red band ($r_{\rm lim} = 17.77$), and its depth is similar to 2dFGRS. The survey will cover a cosmological volume as large as $1[h^{-1}\ \mathrm{Gpc}]^3$. Taking the length scale of $100\ h^{-1}$ Mpc as a tentative scale of homogeneity, this is approximately thousand times a representative volume of the Universe. Hence when fully complete, SDSS can be fruitfully used to evaluate the cosmic variance for many statistics associated with LSS. The data are already large enough to permit accurate estimation of power spectrum[49], 3-point correlation function[50], and of cosmological parameters[51]. Complementary studies of LSS in SDSS have been undertaken using MFs by[44] and using the minimal spanning tree-formalism by[46]. We shall mention later some results from an MF-based study of LSS in mock catalogues of SDSS[52].

In order to make full use of the information encoded in the observed LSS, it is essential to understand the relationship between the number density of galaxies and the underlying mass density field. Since 1980s, it has been realized that the *visible* density field due to galaxies may be *biased* relative to the mass-density. In order to appreciate the implications that bias may have in the program of confronting theory with observations, below we briefly review the state of our knowledge in this regard. First we shall discuss the theoretical modelling of LSS using numerical simulations.

III. N-body simulations

Strong evidence exists to suggest that dark matter is ~10 times more abundant than visible baryonic matter. Dark matter is usually assumed to be collisionless and to interact only through gravitation. Due to its relative abundance dark matter is clearly more important than baryons from a dynamical point of view. The best way to understand the evolution of a gravitating system consisting of N dark matter particles would be to *actually* solve for the force acting on a given particle due to the remaining particles. In a rather different form, by solving for the Poisson's equation at every stage of evolution of the matter-field, this is precisely the goal which modern versions of N-body simulation codes achieve. (For an excellent introduction to N-body simulations, see refs 53–56).

Simulations allow us to probe the strictly nonlinear regime of the dark matter density field, and can provide us with detailed predictions of the *dark side* of the Universe at today's epoch. (The resulting dark matter distribution can be probed through observations of gravitational lensing – an area of great future promise.) The large scale structure distribution in dark matter could very well resemble that shown in Figure 2. We note from this f gure, how the matter condenses along filamentary/sheet-like ridges between clusters and how voids get progressively more evacuated with the passage of time. Due to web-like and frothy appearance, LSS has been described at various times as being a *Cosmic Web*[57] or as *Cosmic Froth*[13].

During the past two decades the advent of parallel supercomputers has ensured a phenomenal improvement in the computing power available to tackle c(nplex astrophysical processes. Hence, it has become possible to include complex gas-dynamical processes into simulations. Various recipes for cooling, star formation and feedback due to supernovae and galactic winds have been incorporated to mimic the formation and evolution of galaxies in a cosmological setup with simulated volumes ranging from 50 to 100 h^{-1} Mpc[58,59]. This has made it possible to develop mock catalogues of galaxies inspired by a given cosmological model. Relatively less time-consuming semianalytic methods have also been proposed which hope to achieve similar goals while bypassing the complex baryonic physics[12,60–63]. Such mock catalogues can eventually be compared with clustering properties of galaxies derived from redshift surveys with the help of methods reviewed here.

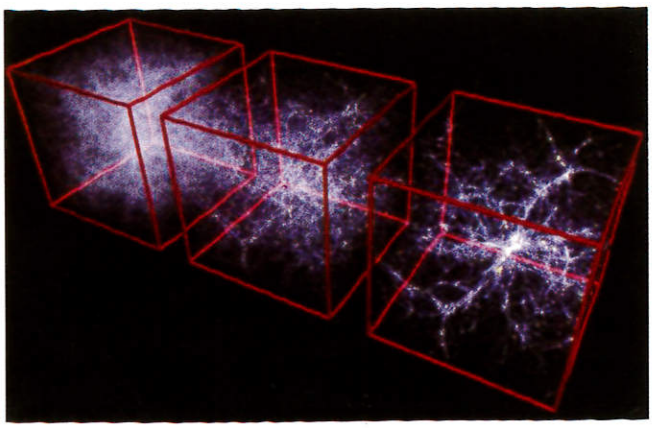

Figure 2. Development of nonlinear structures in a hydrodynamic N-body simulation from $z = 6$ (leftmost cube) to $z = 0$ (rightmost cube) via $z = 2$ (middle cube). A near featureless density field evolves to produce filamentary and sheet-like superclusters which percolate through the box-volume. These are separated by large voids. Due to its web-like appearance, LSS is dubbed as the *Cosmic Web*. [Figure courtesy: Volker Springel; see http://www.mpa-garching.mpg.de/galform/data_vis/index.shtml].

IV. The nature of bias

Above we summarized the theoretical efforts being made to understand the clustering of dark matter. From the predictive point of view, given the fully evolved nonlinear mass-density field (both in the real universe and in N-body simulations), we would like to know the exact locations where galaxies are formed. An important question related to this is whether the clustering properties of the resulting galaxy distribution (quantified in terms of n-point correlation functions or some other alternate statistics) are *the same* as the underlying matter-field, or whether the two are different. The answer to this, clearly depends upon the recipe which we utilize in grafting galaxies on the density field.

During the mid-1990s it was shown that AGNs, IRAS-selected galaxies and galaxies selected in the optical bands cluster differently[64]. Earlier (during the 1980s) it was established that clusters of galaxies are more strongly clustered than galaxies themselves[65]. Recently it was proven that the more luminous galaxies are more strongly clustered[39,66]. Since these different visible tracers of LSS appear to cluster differently, their distributions are clearly biased relative to each other. There is hence, a reason to believe that the visible matter as a whole will be biased relative to the underlying dark matter. This phenomenon is termed as *bias*. Knowledge of the nature of bias will clearly enable us to connect the two distributions – visible and dark; observed and simulated.

In cosmology, it is a conventional practice to quantify the clustering of LSS using the two-point correlation function or, equivalently, the power spectrum. Both these statistics are known to have a power-law form for galaxies[34,67]. However, the correlation function of dark matter was shown to differ from $\xi_{gg}(r)$ in a complicated, scale-dependent fashion for various CDM-based cosmogonies[68–70]. This suggests that the bias *cannot be* scale-independent, i.e. light may not follow the mass in a linear fashion. Recent investigation involving the three-point correlation evaluated for the SDSS dataset implies that galaxy biasing could be quite a complex and nonlinear process; see ref. 50.

Kaiser[65] explained the strong clustering of clusters using the *high peak model*. This model states that a rare, high-

density fluctuation corresponding to a massive object of a given size, collapses sooner if it lies in a region of a larger-scale overdensity. This is because, the small-scale overdensity is *aided* by the surrounding large-scale overdensity so that the collapse-criteria is fulfilled relatively easily for such a halo. This 'helping hand' from the long-wavelength mode means that overdense regions contain an enhanced abundance of massive objects with respect to the mean. Later this model was generalized to incorporate objects of any mass[71,72].

The high-peak model is based on kinematic premises. It traces back the observed clustering of various classes of objects to the in-built structure in the primordial density field. It could fruitfully be applied to study the clustering of galaxies only at high redshift[65,71,73] ($z \simeq 2.5$–3), for the galaxy-scale fluctuations were prone to collapse at such early times. At such high redshifts, galaxies were shown to be highly biased relative to background[73]. Starting from $z \gtrsim 3$, until the present, galaxies have very likely collapsed in all environments. The underlying mass-fluctuations have also grown. Thus, one predicts that the large bias (defined as $b(r) = \sqrt{\xi_g(r)/\xi_{dm}(r)}$ which existed earlier, should asymptotically approach unity on large scales. Lahav *et al.* showed for 2dFGRS data that the bias averaged over all scales is statistically consistent with a scale-independent, large-scale bias of order unity. However, we should see whether this knowledge helps us achieve our original goal of connecting the simulations with the observed Universe, i.e. in populating the mass-density field with galaxies in the correct manner.

A recent, *Halo Model* based approach is a fruitful step in this direction. Here all the complications of galaxy formation are encoded via the halo occupation number: the number of galaxies found above some luminosity threshold in a virialised halo of a given mass . In a nut-shell, the halo model describes nonlinear structures as virialized halos of different mass, placing them in space according to the linear, large-scale density field[74]. This model is analytically tractable, and correctly incorporates the dependence of bias on the 2nd order correlations in the density field. Consequently, it correctly reproduces the observed 2-point correlation function and the power spectrum. Figure 3 compares the exact nonlinear mass-density field with the halo model representation. The halo model is remarkably successful in obtaining correct correlation function, power spectrum, etc. However, as shown in Figure 3 it is manifestly different from the fully evolved density field. This brings out an important aspect relevant to further study of LSS: the low-order correlations cannot be relied upon to provide us with a complete picture of clustering. In addition to the large halos of clusters and groups, today the rich filamentary/sheet like patterns in matter density field are also populated by galaxies[75], and one must theoretically understand the emergence of halos along filaments/sheets in order to complete the picture. In fact, the following experiment can be carried out: the galaxies can be *selected* in $z = 0$ mass-density fields from two rival cosmogonies, so that the observed galaxy–galaxy correlation function $\xi_{gg}(r)$ and the power spectrum P_k are reproduced. In comparing our predictions with the LSS-data, if we confine ourselves merely to two-point correlation function, such an algorithm would validate *both* models to be good representatives of our Universe. Evidently this is not true. Clustering properties of galaxies in the two models would of course be different from each other (and perhaps even from LSS in the Universe). This serves to establish that the large-scale distribution of galaxies *cannot* solely be described by the two-point correlation function alone[76]. The resulting galaxy distributions can be shown to differ from each other by employing statistics which are sensitive to higher order correlation functions. Following this, we arrive at an important conclusion: the two distributions of galaxies *must be* compared with the help of higher order correlation functions or statistics derived therefrom.

To summarize, although the *high peak model*[65] gives us promising results for galaxies assembling at high redshift and for Abell clusters at present, the predictive power of *halo model* in populating mass density field with galaxies is clearly inadequate (in fact, due to a variety of reasons, the bias could even be stochastic; see ref. 77). In particular, for a combination of a cosmological model and a biasing scheme to be proven correct, it is *not sufficient* that the two explain the two-point correlation function for galaxies *alone*. In order to effectively compare the theoretical predictions for galaxy-distribution with the data from redshift surveys, one should include information from higher order correlation functions, and/or alternate statistics which can effectively quantify the complex morphology of the galaxy distribution. One such class of diagnostics – the Minkowski functionals – is introduced in the next section.

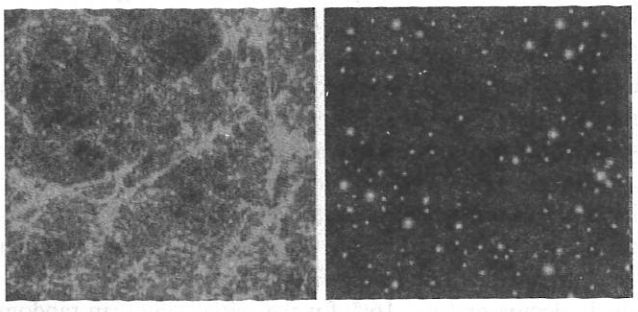

Figure 3. *Left panel:* Simulated dark matter distribution which shows complex morphology. *Right panel:* Halo Model corresponding to density field in the left panel, which successfully reproduces galaxy–galaxy correlation function and the power spectrum. The two are manifestly different, which points out the role which higher order correlation functions play in defining the cosmic structure. Figure courtesy ref. 74.

V. Minkowski functionals

Minkowski functionals in cosmology

It is well known that, in contrast to a Gaussian random field (GRF), a fully evolved nonlinear density field cannot be

fully quantified in terms of its two-point correlation function; the latter being simply the lowest and first of an infinite hierarchy of correlation functions describing the galaxy distribution. Furthermore, the bias in galaxies, *vis-à-vis* dark matter can be nontrivial in nature, and can give rise to different clustering properties of galaxies and the underlying mass density field. Consequently, the correlation function $\xi_{gg}(r)$ determined either for galaxies or for dark matter does not on its own validate a given cosmological model unless support is provided by other, statistically independent measures of clustering. It therefore becomes a challenge to compare the theoretical predictions for LSS with real observational data.

If we knew all n-point correlation functions, we would have a complete description of the galaxy clustering process. However, estimating $\xi_{gg}(r)$ for a sample of N galaxies requires knowing all pairs of galaxies in this sample, whereas calculating the 3-point function implies taking all triplets. The amount of computation escalates rapidly with n – the order in the n-point correlation function – especially for a large value of N. Besides, it is difficult to extract intuitively useful information from these statistics. The present article focuses on another class of statistics which complement the correlation function approach and which have the advantage of providing a physically appealing interpretation for the evolving density field. These are the *Minkowski functionals*.

For an excursion set involving particles embedded in n-dimensions, the Minkowski functionals (MFs) are defined on an $(n-1)$-dimensional hypersurface. There are $n+1$ MFs in n dimensions. In this article, we shall be concerned with the 3-dimensional distribution of dark matter and/or galaxies. Hence, MFs will be defined on a 2-dimensional surface and they will reflect the physical properties of this surface (in the given instance, an isodensity contour referring to a supercluster of galaxies or a void). For a given surface, the four MFs[10,78,79] are, respectively.

1. Volume V,
2. Surface area S,
3. Integrated Mean Curvature C,

$$C = \frac{1}{2}\oint\left(\frac{1}{R_1}+\frac{1}{R_2}\right)\mathrm{d}S. \tag{1}$$

R_1 and R_2 are the two principal radii of curvature of the surface in a given local neighbourhood and the integral is taken over the entire closed surface.

4. Integrated Gaussian Curvature (or Euler characteristic) χ,

$$\chi = \frac{1}{2\pi}\oint\left(\frac{1}{R_1R_2}\right)\mathrm{d}S. \tag{2}$$

A related quantity which is more popular in cosmology is the genus G. The genus is related to χ by

$$G = 1-\frac{\chi}{2}. \tag{3}$$

The 3-dimensional genus of an object is a topological invariant. It can be interpreted in terms of the connectivity of the surface. In simple terms it can be viewed as the number of independent cuts which one can make to the surface *without* breaking it into two separate pieces. A torus can be cut *once* and yet remain in one piece. Hence, its genus is 1. However, a sphere would separate into two pieces if a cut were made. Hence, its genus is 0. Two objects with the same value of genus are topologically similar: one can be obtained by continuously deforming the other. Thus, a sphere and a cube are topologically equivalent. More concretely, genus of an object is the number of handles that the object has, in excess of the number of holes which it encloses (e.g., see ref. 80). Thus, G = [number of handles to the surface] – [number of holes enclosed by the surface].

According to this definition, a sphere has no handle, a torus has a single handle (equivalent to a sphere with one handle), and a pretzel has two handles. Introducing a hole or a bubble *inside* the surface *reduces* its genus by 1, whereas adding a handle to the surface *increases* its genus by 1.

An attractive feature of MFs is that they depend upon the entire hierarchy of correlation functions[81]. Thus, MFs can indirectly help us capture the effect of n-point functions. Higher order n-point functions gradually become important as a primordial, featureless GRF evolves to develop nonlinear structures. This reflects in the behaviour of MFs for the corresponding density field. MFs are additive in nature, i.e. they can be studied for individual objects (say, clusters or superclusters defined using some prescription) as well as for the entire ensemble of such objects.

Of the 4 MFs listed above, the genus G was already known to the cosmology community[5,6,82]. Tools like percolation analysis which can also be related to MFs, had earlier been introduced in cosmology[83]. There were strong reasons why this was the case.

It is well known that a system evolving under gravitational instability becomes progressively more non-Gaussian. (The hypothesis of a primordial spectrum of density perturbations which is distributed in the manner of a Gaussian random field is supported by observations of the cosmic microwave background carried out by WMAP and other experiments.) For CDM-like cosmological models this results in more matter being concentrated in filamentary and pancake-like distributions whose coherence scale evolves with time. As a result, the filling fraction at percolation progressively decreases from 16% for the initial Gaussian random field to ~5% for LCDM[84–87]. This implies that more matter is being transferred to regions which occupy a progressively smaller amount of space and are also spatially anisotropic and therefore percolate easily. Figuratively this corresponds to the prominance acquired by superclusters

of galaxies over time – see Figure 2. This theoretical picture which described the emergence of structure in an initially featureless medium agreed well with what observations were telling us about our universe. The emerging picture of LSS taken from redshift surveys shows us that galaxies are distributed preferentially along filaments and sheets which seem to encircle vast voids. The distribution of galaxies does not occupy much volume, and because of the presence of superclusters, percolates easily. Percolation theory can therefore be used to describe an important quality of the observed galaxy distribution – its connectivity. At the percolation threshold one can visit the far corners of the cosmic web by following filamentary/sheet-like overdensities. The cosmic web is therefore a connected structure and this is clearly revealed in surveys such as Figure 2. Connectivity is of course a mathematical notion, and it is possible to precisely quantify connectivity of an object or a system by evaluating its percolation properties and its genus. Doroshkevich[82] gave analytic formula for the genus of a GRF which showed that at the median density threshold, GRF would exhibit a sponge-like topology[82]. Simulations of LSS within CDM-cosmogonies provided similar visual impression when smoothed on sufficiently large scales. Like percolation, the genus too has come to be considered as a useful probe of the initial Gaussianity (as well as the final non-Gaussianity) of the cosmic density field; see for instance refs 6, 25, 85.

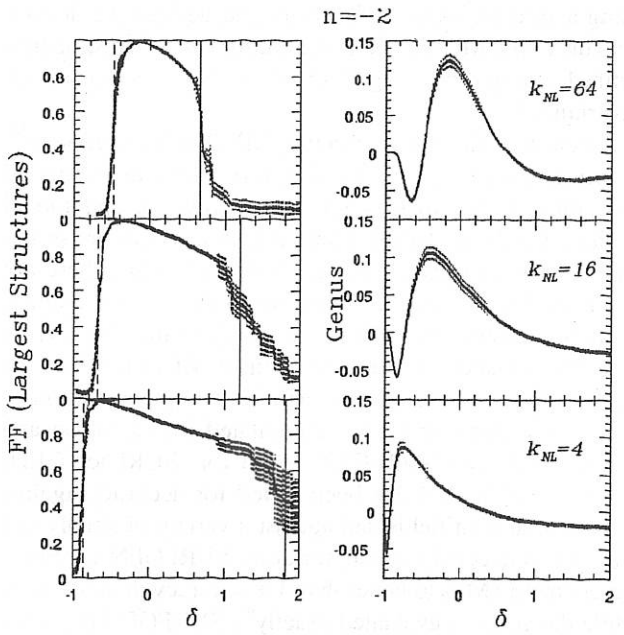

Figure 4. Percolation (left panels) and genus (right panels) curves are plotted as functions of the density contrast δ for scale free models of gravitational clustering with the perturbation spectral index $n = -2$. Solid and dashed curves in the left panels correspond to the percolation curve for the largest cluster and void respectively. Vertical solid/dashed lines mark the threshold describing percolation between opposite faces of the cube for clusters/voids respectively. Figure courtesy of ref. 85.

An important indicator of non-Gaussianity in a distribution is the *percolation curve*. The percolation curve describes the *volume fraction* (or filling factor – FF) in the largest structure (cluster/void) as a function of the density contrast threshold[88] δ. Salient features of the percolation curve are illustrated in Figure 4 in which density fields (with an initial power spectrum $P(k) \sim k^{-2}$) evolve from an epoch when the scale of nonlinearity is $k_{NL} = 64k_f$, $16k_f$, $4k_f$, (k_f is the fundamental mode corresponding to the box-size of our N-body simulation). In Figure 4 percolation curves for clusters (thick solid lines) and voids (thick dashed lines) are shown as functions of the density contrast δ. Starting from a high density threshold (small *FF*) we find several isolated clusters (corresponding to peaks of the density field). Lowering the density threshold further results in the *merger* of clusters leading to a rapid growth in the percolation curve and to the onset of percolation. A further lowering of the threshold to very small values results in the merger of almost all clusters so that $FF \rightarrow 1$. An identical procedure followed for underdense regions by gradually increasing the density contrast threshold (increasing *FF*) results in a similar functional form for the volume fraction in the largest void. In our samples the largest cluster percolates between opposite faces of the cube when its filling factor is about half the total *FF*. In most cases percolation also coincides with the highest jump in the volume of the largest cluster which was used[89] as a working definition of the percolation threshold.

The solid/dashed vertical line in Figure 4 represents the density contrast threshold δ_C below/above which clusters/voids percolate. We note that *both* clusters and voids percolate over a range of overlapping density contrasts – a feature that is only possible in three or more dimensions – and corresponds to what is commonly called a *sponge* topology for the density distribution.

As the simulation evolves δ_C increases monotonically for $n = -2$, as structures form and align on increasingly larger scales. From Figure 4 we see that voids find it easier to percolate as the simulation evolves, as a result the range in densities when both phases percolate initially increases, enhancing the extent of *sponge-like topology* in the distribution.

The right panel in Figure 4 shows the evolution of the genus curve for the same simulation. The top panel corresponding to an early epoch shows the genus curve largely retaining its bell-shaped form which it has for the primordial Gaussian random field[90,91] $G(\nu) = A(1-\nu^2)\exp(-\nu^2/2)$. During later epochs the density distribution grows progressively more non-Gaussian and this is rejected in the change in both shape as well as amplitude of the genus curve in Figure 4.

Although a study of the genus gives us a useful handle on the connectivity of a density field, we still lack information about the morphology of LSS. For example, if the domain of information is solely restricted to knowing genus, a filament with one handle would be considered to be

identical to a sphere with one handle or a pancake with one handle. Therefore, to gain objective insights into the nature of the supercluster-void network, we must complement genus of an object with quantities which in some way, characterize geometry of that object. This is precisely the role played by the first three MFs, Volume V, Surface Area S and Integrated Mean Curvature C. These MFs change when local deformations are applied to the surface, and hence, these may be useful to glean information about the typical size of an object. The final goal here would be to use this geometric information and give an objective meaning to generic sizes and shapes of superclusters and voids belonging to LSS. In this article we will show that MFs can be utilized to characterize the geometry and topology of the cosmic density field, and eventually can be employed to get information about the morphology of the supercluster-void network in LSS. Because MFs depend on the full hierarchy of correlation functions, we can use them to compare and distinguish two rival cosmological models. Recently, Matsubara[80] derived semi-analytical expressions of MFs for weakly nonlinear cosmic density field by using a second-order perturbative formalism. The results obtained by Matsubara can be used in conjunction with the methods reviewed here to obtain insight into the dynamics of the clustering process of a density field smoothed on sufficiently large scales. The role of bias on large scales can also be directly probed using such methods.

Morphology of LSS with Shapefinders

Sahni *et al.*[92] showed that the MFs could be used to evaluate both the size as well as the shape of a three-dimensional object such as a supercluster or a void. The size is given in terms of three Shapefinders which have dimensions of length. These are defined as ratios of MFs, and are conveniently termed as Length ($\mathcal{L}$), Breadth ($\mathcal{B}$) and Thickness ($\mathcal{T}$).

$$\mathcal{T} = \frac{3\times V}{S} \tag{4}$$

$$\mathcal{B} = \frac{S}{C} \tag{5}$$

$$\mathcal{L} = \frac{C}{4\pi G}. \tag{6}$$

The three Shapefinders are defined using spherical normalization, so that for a sphere of radius R, $\mathcal{T} = \mathcal{B} = \mathcal{L} = R$.

The above measures quantify the size of the object in question. To further quantify the shape of these objects, two dimensionless Shapefinders have been defined as follows. These are Planarity ($\mathcal{P}$) and Filamentarity ($\mathcal{F}$).

$$\mathcal{P} = \frac{\mathcal{B} - \mathcal{T}}{\mathcal{B} + \mathcal{T}} \tag{7}$$

$$\mathcal{F} = \frac{\mathcal{L} - \mathcal{B}}{\mathcal{L} + \mathcal{B}}. \tag{8}$$

Sahni *et al.*[92] showed that ($\mathcal{P}$, $\mathcal{F}$) quantify the shape of both simple as well as complicated objects. Thus, for an oblate ellipsoid its planarity is relatively large and its filamentarity is small. For a prolate spheroid, the reverse is true. The statistics respond monotonically to deformations of these surfaces.

If we include the genus then, the triplet of numbers ($\mathcal{P}$, $\mathcal{F}$, $\mathcal{G}$), can define a three-dimensional *Shape-Space* which may be used to represent the distribution of shapes of superclusters and their topologies in a given cosmological density field. Note that the inclusion of genus is significant, since most superclusters at moderate density thresholds can have a spongy texture with several loops and/or branches emanating from a central body (see Figure 11 for an illustration).

Methods of estimation

In cosmology the distribution of galaxies in space can frequently be regarded as being an example of a point process. There have been several attempts of evaluating MFs for a given point process. For example, the Boolean grain model[10,81] would decorate the input set of points with spheres of varying radii, and would study the morphology of the structures which contain overlapping spheres. According to another approach[11], one smooths the distribution of points using a suitable window function, and defines the density field on a grid. MFs can be evaluated for connected structures defined on a grid by using Crofton's formulae or Koenderink Invariants[11].

A *new* approach for determining MFs has been reported[93]. Rather than working with a grid-approximation for connected structures, these workers go a step forward, and model surfaces for these objects. Minkowski functionals are evaluated for the isodensity surfaces *online*, i.e. while these are being modelled using an elaborate surface triangulation technique. Sheth *et al.*[93] describe the algorithms for surface construction and for the evaluation of MFs on these surfaces. Sheth[94] discusses the method of implementation of these algorithms which has culminated into a robust and accurate software SURFGEN (short for 'SURFace GENerator'). SURFGEN has been tested for accuracy against Gaussian random fields and against a variety of simply and multiply connected eikonal surfaces. SURFGEN evaluates the geometric MFs to better than 1% accuracy in most cases, while the genus is evaluated exactly[93]. SURFGEN has since been employed to study geometry, topology and morphology of LSS *both* in dark matter and galaxies within a variety of cosmogonies. (SURFGEN could also be relevant in areas of science where the properties of surfaces becomes important such as in medical imaging (tomography) and in condensed matter physics.)

In the next section we shall discuss a set of new results obtained by applying MF-based techniques to LSS-data.

VI. Results

Morphology of large scale structure using shapefinders

Given a cosmic density field, MFs can be evaluated for a supercluster (void) defined as a connected overdense (underdense) region above (below) certain physically motivated threshold of density. SURFGEN-like technique would isolate such a structure, and would model its isodensity contour in order to precisely estimate its MFs. MFs referring to a given structure are called *partial MFs*. By knowing partial MFs, one can further measure shape and size of the concerned structure using Shapefinders. Further, MFs are additive in nature. Hence, at any given threshold of density, *global MFs* pertaining to the entire density field can be obtained by summing over *partial MFs* of the constituent overdense/underdense objects. While partial MFs are useful to measure shapes and sizes of individual superclusters and voids, global MFs can be employed (1) to discriminate between models, (2) to probe the effect of nonlinear gravitational evolution on the density field and (3) to capture the nontrivial effects of bias on the structure of the cosmic web. We shall review here some of the promising results obtained in these directions.

Discriminating between rival models of LSS

The growth-rate of density perturbations differs between various cosmogonies. As a result, the clustering properties of the matter-distribution also turn out to be different. A serius candidate statistic of LSS should be able to capture such differences and discriminate between rival cosmological models.

The three upper panels of Figure 5 show clustered distribution of dark matter at $z = 0$ in three rival cosmological models – SCDM, τCDM and ΛCDM respectively, simulated by Virgo group[70]. The three models are simulated such that the observed abundance of rich clusters of galaxies can be reproduced in *all* of them. SURFGEN was employed to quantify the visually apparent differences between these models. Global MFs were evaluated at a set of density levels, and were studied with respect to FF_V, the fractional *overdense* volume[93]. The lower three panels of Figure 5 show variations in three global MFs – area, curvature and genus – as density level is progressively brought down. Notice that MFs for SCDM (dashed lines) show the largest amplitudes, those for ΛCDM (solid lines) show the smallest amplitudes with τCDM curves (dot-dashed lines) falling between the two. Since the number of rich clusters and their relative spatial distribution in all the models is the same, this difference can be attributed to difference in matter-distribution along sheets and/or filaments connecting the clusters[93,94]. Global MFs are thus noted to be remarkable at discriminating between models (also see ref. 79).

In Section IV we noted that the distribution of galaxies cannot be fully quantified using its 2-point correlation function $\xi_{gg}(r)$ alone. However, the galaxy distribution predicted by a given cosmological model must *at least* reproduce the observed $\xi_{gg}(r)$. Cole *et al.*[96] generated *mock* catalogues of galaxies in a variety of cosmogonies subject to the above constraint. Thus, galaxies were selected from the simulated $z = 0$ mass-distributions so that the observed $\xi_{gg}(r)$ is reproduced. The catalogues were further used to construct *mock* 2dFGRS and SDSS redshift surveys in such models. Since MFs depend on the hierarchy of correlation functions, these can be relied upon in distinguishing the *galaxy distributions* resulting from different models. Sheth[94] made such a comparison between mock SDSS catalogues of galaxies from ΛCDM and τCDM models by evaluating global MFs of density fields smoothed with a Gaussian window 6 h^{-1} Mpc wide. Figure 6 shows the global mean curvature and genus for the two models evaluated using SURFGEN. Note that the MF-amplitudes in ΛCDM are *larger* than those for τCDM model, and that the two models are successfully distinguished from each other using MFs[52], *even* in the case when the distributions share the same $\xi_{gg}(r)$. (The τCDM curves are averaged over 10 realizations made available[96].) MFs can thus be reliably adopted to compare mock catalogues with the observed LSS in redshift surveys.

MFs and the scale-dependent bias

We reviewed the status of our knowledge about bias in Section IV. We raised an important question as to whether the clustering properties of dark matter and the *biased* galaxy distribution could be different or otherwise. This question was investigated[52] by employing the mock SDSS catalogues of galaxies in ΛCDM and τCDM models. The clustering properties of galaxies were measured in terms of MFs.

As noted in refs 6, 97, the amplitude of the genus-curve drops as the N-body system develops phase correlations. Given two density fields, the system with larger genus-amplitude shows many more tunnels/voids which are, therefore, smaller in size. As time progresses, the voids are expected to expand and merge, leading to a drop in the genus-amplitude, while the phase correlations continue to grow. With this simple model, one could correlate the amount of clustering with the relative *smallness* of the amplitude of genus, and therefore, of the area and the mean curvature. Going by this reasoning, we may infer from Figure 6 that the ΛCDM galaxy-distribution is relatively *more* porous and would show *less* coherence on large scales compared to that due to τCDM. As illustrated in Figure 5 (see ref. 93), the dark matter distribution of τCDM model due to Virgo group shows considerably larger amplitudes for the MFs compared

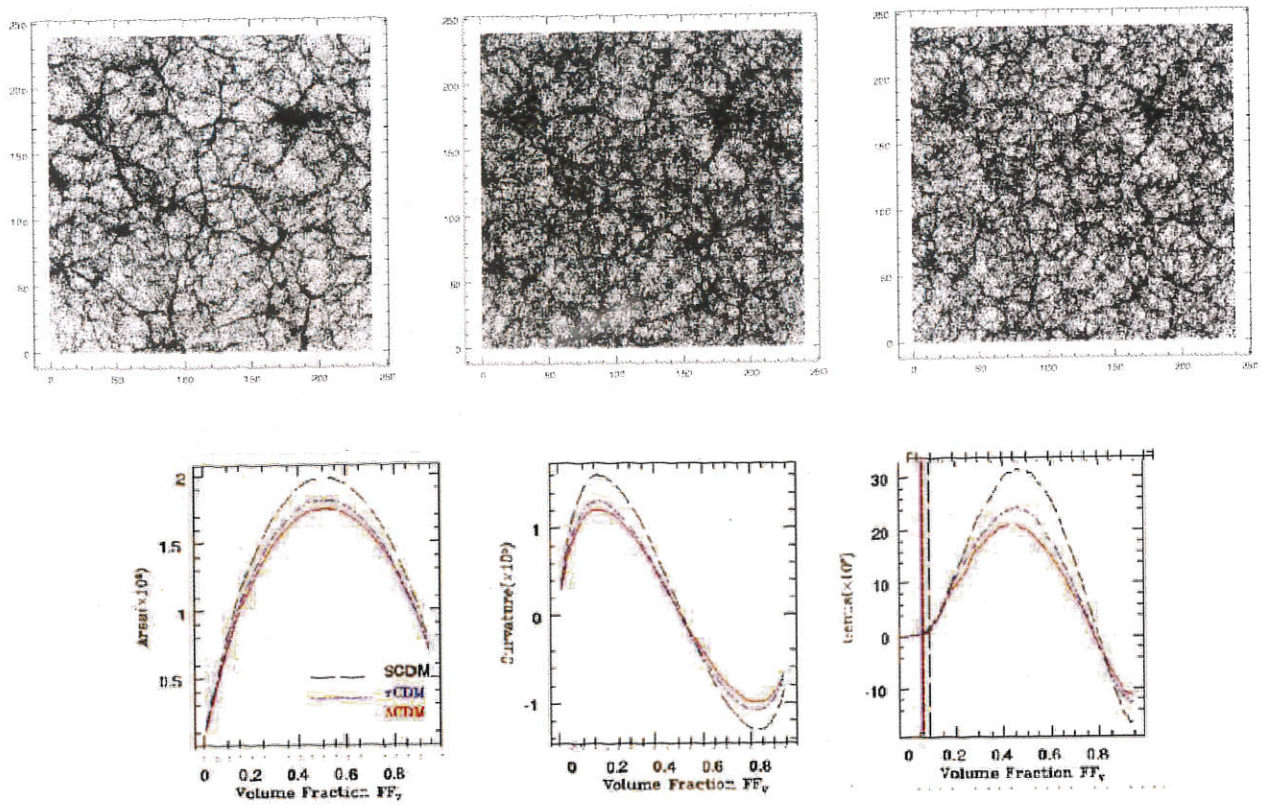

Figure 5. Three global MFs for ΛCDM (solid lines), τCDM (dot-dashed lines) and SCDM (dashed lines) are shown as functions of the overdense volume fraction FF_V. The three models have appreciably different morphology and hence can be distinguished from one another on the basis of MFs. For further discussion, please refer to the text. Figure courtesy of ref. 93.

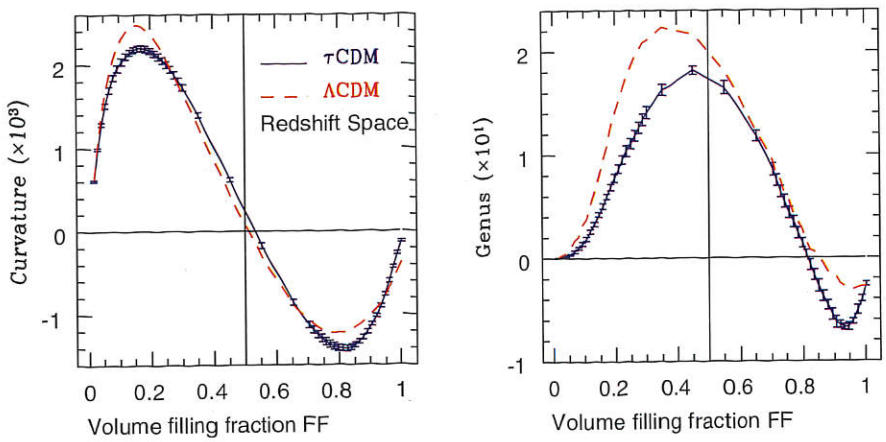

Figure 6. Two global MFs – curvature and genus – are evaluated at 50 density levels at a common set of volume filling fractions and are studied here with respect to FF_V. The values are normalized to the volume of $[100\ h^{-1}\ Mpc]^3$. The global MFs for τCDM model are averaged over 10 realizations (solid lines) and the error bars represent 1σ deviation. Assuming the same level of accuracy for ΛCDM, we may conclude that these MFs with volume parameterization can indeed *clearly* distinguish τCDM from ΛCDM. Figure courtesy of ref. 52.

to the ΛCDM model, whereas we find the reverse trend in the MFs of the *galaxy distributions* due to the same two models. Evidently biasing appears to be a source of this effect.

To establish this effect more firmly, the τCDM and ΛCDM dark matter Virgo simulations were analysed by adopting the same resolution ($\ell_g = 3.5\ h^{-1}$ Mpc) and the smoothing scale ($\lambda_s = 6\ h^{-1}$ Mpc) as utilized in the analysis of mock catalogues. So as not to introduce any bias due to redshift space distortions, global MFs of the galaxy catalogues were computed in *real space*. The effect of biasing is most dramatically seen in

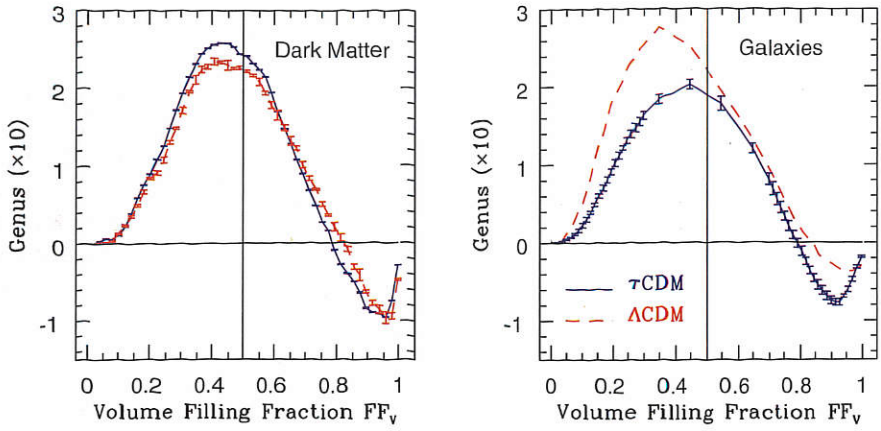

Figure 7. *Left panel*: Global genus curves of ΛCDM and τCDM *matter distributions* are shown. The values refer to a volume of $[100\ h^{-1}\ \mathrm{Mpc}]^3$. The 1σ error-bars are due to 5 realizations of both the models, each with 25% of the total number of particles. Calculations are carried out on a grid of resolution $3.5\ h^{-1}$ Mpc after smoothing the density fields with $6\ h^{-1}$ Mpc Gaussian kernel. *Right panel*: The galaxy catalogues due to the two models are analysed in *real* space under identical conditions. The figure illustrates the phenomenon of *phase-reversal*. For more details, please refer to the text. Figure courtesy of ref. 52.

the global genus-curve of the dark matter and galaxies (see Figure 7). We notice that the amplitudes $\mathcal{G}_{\mathrm{DM}}^{\Lambda} < \mathcal{G}_{\mathrm{DM}}^{\tau}$, whereas $\mathcal{G}_{\mathrm{G}}^{\Lambda} > \mathcal{G}_{\mathrm{G}}^{\tau}$. Here subscripts DM and G stand for 'dark matter' and 'galaxies', respectively. We may conclude from here that the matter distribution in ΛCDM is less porous than that in τCDM, but this trend reverses as we investigate biased distributions of galaxies due to the same two models: the galaxy distribution due to ΛCDM appears to be *more* porous than that due to τCDM. This could perhaps mean that a simple scheme of scale-dependent biasing leads to different degree of phase-correlations in the galaxy distribution as compared to the underlying mass distribution. This can be dubbed as a phenomenon of *phase-mismatch* between the two distributions – visible and dark[52]. This is supported by the fact that[96] require large anti-bias in high density regions of the ΛCDM model.

To conclude, the study of the global MFs reveals that, local, density-dependent bias could lead to an apparent *phase-mismatch* or different phase-correlations among the dark matter and the galaxies. It therefore becomes important to include realistic treatment of bias before comparing theoretical predictions about LSS with redshift surveys. We noted earlier that morphology of LSS appears web-like with frothy voids separating sheet-like and/or filamentary superclusters. The rich texture of LSS is a culmination of two interconnected factors: (1) strong non-Gaussianity induced in cosmic density field due to gravitational dynamics and (2) biased formation of galaxies relative to the underlying mass. Superclusters are thus strongly non-Gaussian, baryonic structures. Unlike galaxies and their clusters, superclusters are as yet unrelaxed and dynamically evolving; their morphological properties may be expected to be sensitive to the cosmological parameters of the Universe. It is hence important to quantify their shapes and sizes and relate them with physically relevant other quantities such as the mass enclosed or volume occupied by them. Similarly, voids *too* can be studied, and the combined properties of the supercluster-void network could provide us with yet another, independent check on the cosmological parameters. Since modern redshift surveys are deep and large enough, this may perhaps be the most effective method to test theoretical predictions.

Recently, a comprehensive morphological study of the supercluster-void network in ΛCDM cosmogony was carried out[87]. Dark matter distribution at $z = 0$ was smoothed with $5\ h^{-1}$ Mpc Gaussian window to produce a cosmic density field sampled over a box of size $239.5\ h^{-1}$ Mpc. Superclusters and voids were defined as overdense and underdense connected regions respectively, and their MFs and Shapefinders were measured using SURFGEN.

Superclusters in ΛCDM model were found to percolate through the Universe while occupying fractional overdense volume as small as 7%[87]. This points to the connectedness of the overdense regions in the universe. From their tendency to effectively percolate the volume, it is apparent that superclusters must show strong departure from sphericity. Voids, on the other hand, occupy as large a fractional volume as 22% before the underdense regions percolate[87]. Evidently, the morphology of voids would be markedly different from superclusters. In shape, they are likely to be more isotropic.

Shandarin *et al.*[87] further supported the above inferences by studying the sizes of superclusters and voids with respect to the enclosed mass and occupied volume respectively. Figure 8 shows the results. Notice that all three sizes of the structures show significant correlation with mass enclosed

(for superclusters) and volume occupied (for voids): the larger the mass, the larger the size of SC and more the volume, the larger is the size of the void, as expected. The solid lines show the radius of a sphere having the same volume as a given object ($R = (3V/4\pi)^{1/3}$). The thickness and breadth approximately double their value and length grows by over an order of magnitude when the mass increases from about $10^{14.5}\ M_\odot$ to $10^{16.5}\ M_\odot$. Both the thickness and breadth are considerably smaller than the radius R of a sphere having similar volume for large superclusters ($M \gtrsim 10^{15}\ M_\odot$). On the other hand the length is considerably greater than R. This is a clear manifestation of the anisotropy of large-scale superclusters (see Figure 11). Voids demonstrate similar correlations with their volumes as shown in the right panels of Figure 8. Voids, especially large ones, are also anisotropic, e.g. see ref. 87 for illustrations. It was shown for example, that *both* superclusters and voids show significantly larger filamentarity in proportion to the mass enclosed and volume occupied, respectively. Superclusters show negligible planarity whereas, voids could be relatively more planar[87].

In their visual appearance, filaments stand out remarkably well in our impression of LSS. However, the occasional occurrence of vast, sheet-like superclusters such as the *SDSS Great Wall* (which is almost twice as large as the CfA Great Wall[49,98]; see Figure 9) raises questions as to how significant are sheets with respect to filaments on large scales in the actual Universe. In future, if the sheet-like superclusters in redshift surveys are confirmed and their counterpart structures *not* found in N-body simulations, it may present a clear dichotomy between theory and observations.

To investigate this question in the context of simulations, Sheth *et al.*[93] studied the correlation in the Shapefinders of ΛCDM superclusters identified at percolation. Figure 10 is a scatter plot of Shapefinders $\mathcal{T}$, $\mathcal{B}$, $\mathcal{L}$ for clusters in ΛCDM defined at the percolation threshold. The strong correlation between $\mathcal{T}$ and $\mathcal{B}$ in the left panel indicates that two (of three) dimensions defining any given cluster assume similar values and are of the same order as the correlation length. ($\mathcal{T} \simeq \mathcal{B} \simeq 5\ h^{-1}$ Mpc for the largest superclusters.) The clustering of objects near $\mathcal{T} \simeq \mathcal{B}$ ($\mathcal{P} \simeq 0$) in this panel suggests that the superclusters are either quasi-spherical or filamentary. The superclusters in simulations thus *do not* appear to be planar. The scatter plot between $\mathcal{B}$ & $\mathcal{L}$ in the right panel of Figure 10 breaks the degeneracy between spheres and filaments. The mass-dependence of morphology is highlighted in this panel in which larger dots denote more massive objects. This figure clearly reveals that more massive clusters/superclusters are, as a rule, also more filamentary, while smaller, less massive objects, are more nearly spherical[93].

For Shapefinder-like statistics, it is important that the measured properties of an object also conform with its visual appearance. Figure 11 shows the largest percolating supercluster in the ΛCDM cosmogony[93] for which $\mathcal{F} = 0.81$. Note that the supercluster is indeed highly filamentary in agreement with the measured, high value of $\mathcal{F}$.

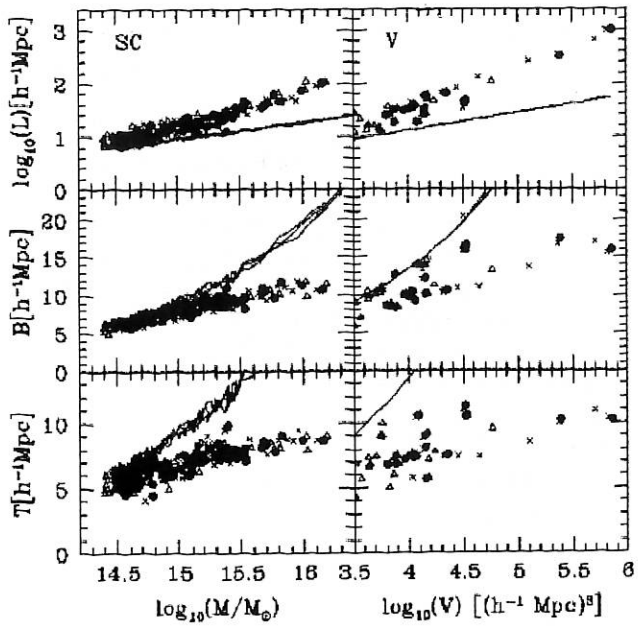

Figure 8. The length, breadth, and thickness versus mass for superclusters and versus volume for voids at percolation. Solid circles show the relation at percolation thresholds: $FF_C = 0.07$ for superclusters and $FF_V = 0.22$ for voids. Crosses show the parameters before percolation ($FF_C = 0.06$ for superclusters and $FF_V = 0.21$ for voids) and empty triangles after percolation ($FF_C = 0.08$ for superclusters and $FF_V = 0.23$ for voids). Solid lines show the radius of the sphere having the same volume as the corresponding object. Note the logarithmic scale used for the length. Three lines correspond to three different thresholds. Figure courtesy of ref. 87.

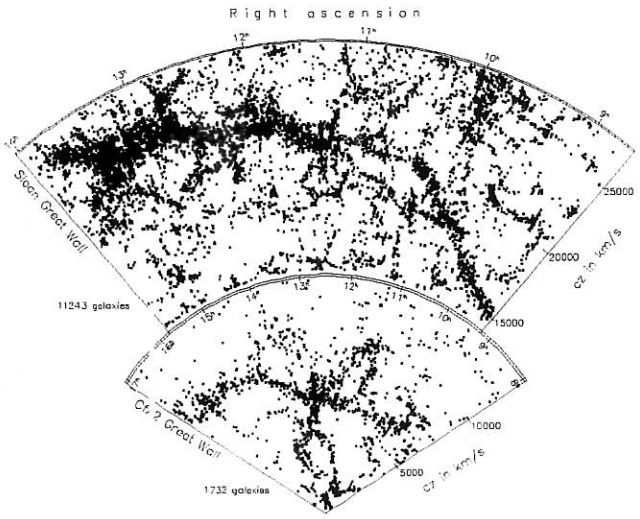

Figure 9. The *Sloan Great Wall* situated at $z \simeq 0.08$ is shown in the upper panel of the figure. For comparison, a relatively nearer structure of the CfA slice is also shown. The linear extent of the Sloan Great Wall is ~420 h^{-1} Mpc, which is about twice larger than the CfA great wall. Figure courtesy of ref. 98.

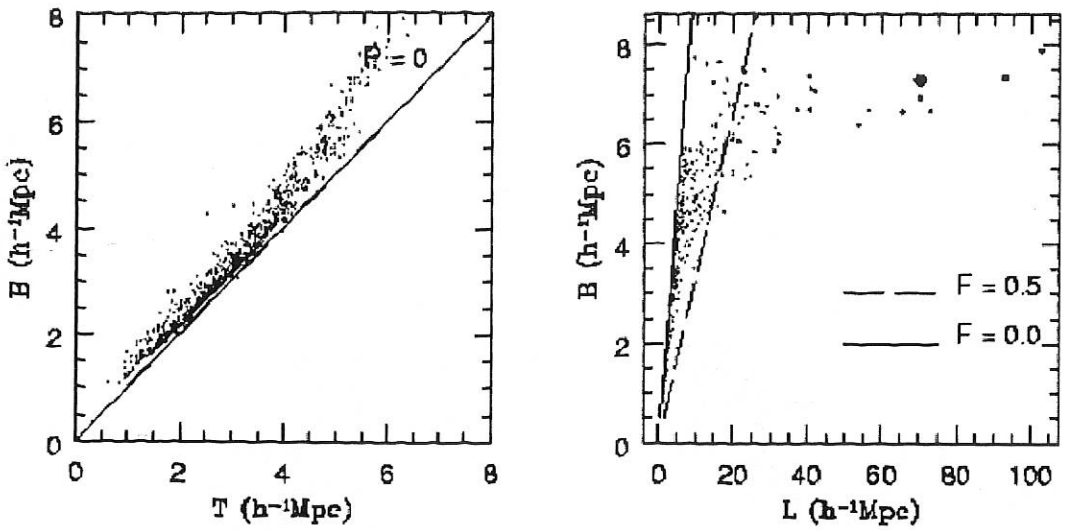

Figure 10. Scatter plot for the pair of Shapefinders $\mathcal{T}$, $\mathcal{B}$ (left panel) and $\mathcal{B}$, $\mathcal{L}$ (right panel) defining the morphology of clusters/superclusters in the ΛCDM model. The strong correlation between $\mathcal{T}$ and $\mathcal{B}$ in the left panel near the line $\mathcal{P} = 0$ indicates that two of the three dimensions defining a cluster are equal and of the same order as the correlation length (≃ few Mpc). Judging from the left panel we find that clusters/superclusters in ΛCDM are either quasi-spherical or filamentary (since both satisfy $\mathcal{T} \simeq \mathcal{B} \Rightarrow \mathcal{P} \simeq 0$). The degeneracy between spheres and filaments is lifted by the right panel which is a mass-weighted scatter plot for the Shapefinders $\mathcal{B}$, $\mathcal{L}$. Each dot in this panel refers to a cluster and its area is proportional to the fraction of mass in that cluster. The concentration of points near the line $\mathcal{F} = 0$ ($\mathcal{B} = \mathcal{L}$) reflects the fact that a large number of smaller clusters are quasi-spherical. The more massive structures, on the other hand, tend to be filamentary and the largest and most massive supercluster has $\mathcal{F} = 0.81$. All objects are determined at the percolation threshold. Figure courtesy of ref. 93.

Table 1. The ten most voluminous superclusters (determined at the percolation threshold) are listed with their associated Minkowski functionals (volume, area, curvature and genus) and Shapefinders $\mathcal{T}$, $\mathcal{B}$, $\mathcal{L}$ in case of the ΛCDM model simulated by Virgo group. The first row describes the percolating supercluster and appears in boldface. It should be noted that the interpretation of $\mathcal{L}$ as the 'linear length' of a supercluster can be misleading for the case of superclusters having a large genus. In this case $\mathcal{L} \times (G + 1)$ provides a more realistic estimate of supercluster length since it allows for its numerous twists and turns

Model	Volume $(h^{-1}\,\mathrm{Mpc})^3$	Area $(h^{-1}\,\mathrm{Mpc})^2$	Curvature $(h^{-1}\,\mathrm{Mpc})$	Genus	Shapefinders (h^{-1} Mpc) $\mathcal{T}$	$\mathcal{B}$	$\mathcal{L}$
ΛCDM	$\mathbf{8.45 \times 10^4}$	$\mathbf{4.5 \times 10^4}$	$\mathbf{6.16 \times 10^3}$	**6**	**5.63**	**7.30**	**70.03**
$\delta = 2.31$	$3.21 \times 1\,)^4$	1.7×10^4	2.33×10^3	1	5.63	7.34	92.74
	2.22×10^4	1.22×10^4	1.76×10^3	1	5.47	6.93	70.00
	2.13×10^4	1.02×10^4	1.29×10^3	0	6.27	7.90	102.7
	1.21×10^4	6.35×10^3	8.78×10^2	0	5.71	7.22	69.95
	1.19×10^4	6.82×10^3	1.02×10^3	1	5.25	6.70	40.53
	1.08×10^4	6.12×10^3	9.16×10^2	0	5.3	6.68	72.90
	1.01×10^4	5.46×10^3	8.19×10^2	0	5.57	6.65	65.24
	8.6×10^3	4.74×10^3	7.07×10^2	0	5.44	7.00	56.33
	8.1×10^3	4.25×10^3	5.71×10^2	1	5.72	7.44	22.74

Earlier we noted results from MF-based study of mock SDSS catalogues due to ΛCDM and τCDM models. Global MFs were shown to successfully discriminate between the two models. Further, we noted the phenomenon of *phase-mismatch* between dark matter and galaxies due to biasing. In this subsection, we noted that superclusters in a typical CDM-cosmogony are generically filamentary. Could we use the *length* of the filamentary superclusters as discriminatory statistic between models? In other words, are the morphological properties of superclusters *indeed* sensitive to the underlying cosmic parameters? The 10 most voluminous superclusters in dark matter Virgo simulations of ΛCDM, SCDM and τCDM models were shown to have considerably different morphological properties. In fact, the percolat-

ing ΛCDM supercluster (shown in Figure 11) was found to be more filamentary and topologically more simpler ($G = 6$) than its counterpart structures in SCDM and τCDM ($G \simeq 20$).

To answer the same question in case of galaxy distributions, Sheth studied the abundance of superclusters larger than certain length $\mathcal{L}$ from within a given large ensemble of superclusters selected from mock SDSS catalogues of ΛCDM and τCDM models. This quantity is referred to as Cummulative Probability Function (CPF). Figure 12 shows the results, where the CPF(L) have been evaluated at $FF_V = 7.8$ per cent (at the onset of percolation) and at $FF_V = 11.7$ per cent (after the percolation). Note that at the onset of percolation, the length of the longest τCDM superclusters could be as large as 90 h^{-1} Mpc, whereas their ΛCDM counterpart structures, which exhibit the same degree of statistical significance, are relatively shorter with $\mathcal{L}_{max} = 5$ h^{-1} Mpc. We conclude that the τCDM superclusters tend to be statistically *longer than* their ΛCDM counterpart structures. The large-scale coherence in the superclusters is attributed to the phase-correlations in the density field; the higher the degree of phase-correlations, the larger the length-scale of coherence. Based on this, and the results reported earlier, we would anticipate the τCDM superclusters to be *longer than* those in ΛCDM. As we can see, the result reported here agrees well with this anticipation. This is a *considerable* success for the ansatz of the Shapefinder quantifying *length*, for as we noted above, it helps us capture the relative effect of phase-correlations among rival models of structure formation.

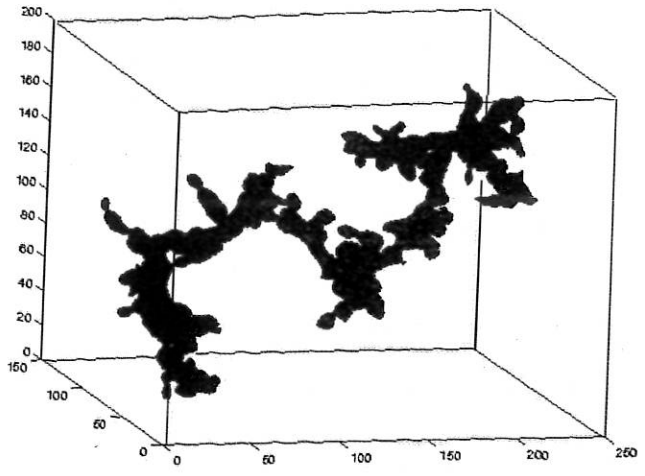

Figure. 11. The largest (percolating) supercluster in ΛCDM. This cluster is selected at the density threshold which marks the onset of percolation ($\delta_{perc} = 2.3$). As demonstrated in the figure, the cluster at this threshold percolates through the entire length of the simulation box. It is important to note that the percolating supercluster occupies only a small fraction of the total volume and its volume fraction (filling factor) is only 0.6%. Our percolating supercluster is a multiply connected and highly filamentary object. Its visual appearance is accurately reflected in the value of the Shapefinder diagnostic assigned to this supercluster: ($\mathcal{T}$, $\mathcal{B}$, $\mathcal{L}$) = (5.63, 7.30, 70.03) h^{-1} Mpc and ($\mathcal{P}$, $\mathcal{F}$, $\mathcal{G}$) = (0.13, 0.81, 6). Figure courtesy of ref. 93.

Applications to redshift surveys

We end this section by providing a brief summary of the results which have emerged out of the analysis of the observed LSS. We further outline the scope of these methods in light of the newly available large datasets.

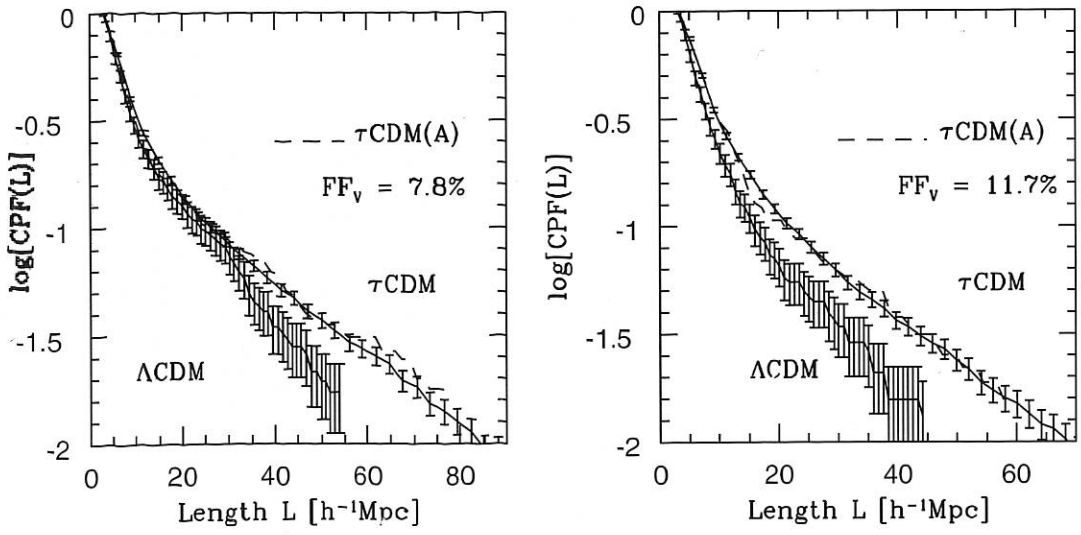

Figure 12. Shown here are the cumulative probability functions of length $\mathcal{L}$ for superclusters of galaxies in τCDM and ΛCDM at two thresholds of density corresponding to $FF_V = 7.8$ per cent (just before the onset of percolation) and $FF_V = 11.4$ per cent (after the onset of percolation). We find the CPFs of the two models to be distinctly different at longer length-scales. The dashed line refers to the CPFs due to the first realization of τCDM which shares the initial set of random numbers with ΛCDM. The largest superclusters of τCDM are systematically larger than those due to ΛCDM. Figure courtesy of ref. 52.

The methods of topological analysis of LSS have been well developed[6] since late 1980s. Hence, most of the investigations complementing the standard correlation function based approach are in the wake of studying the connectivity and topology of LSS. Almost all the topological analyses are carried out to test whether the galaxy-distribution is consistent with the underlying Gaussian, random-phase primordial density field. Topology of LSS revealed both in LCRS data[25] and SDSS data[43] have been found to be consistent with Gaussian, random-phase hypothesis for the primordial density fluctuations. SDSS Early Data Release has been further found to agree well with the predictions of ΛCDM concordance cosmology model[47]. Springel *et al.* reached similar conclusions based on their analysis of the 1.2 Jy redshift survey[97]. Connectivity of LSS of LCRS has been studied by Shanderin and Yess[86]. These authors detected a network of filamentary superclusters in LCRS slices, which they found to be consistent with the gravitational instability paradigm of structure formation.

Recently there has been an upthrust of interest in complementing earlier topological studies with geometry of LSS. Probably the first morphological analysis of LSS was done using 1.2 Jy redshift survey data by Sathyaprakash *et al.*[99] MFs for SDSS early data release (EDR) have been evaluated by Hoyle *et al.*[44] who find the data to be consistent with the clustering predictions of the ΛCDM cosmogony[100]. Basilakos[101] has measured shapes of superclusters in SDSS (EDR) and finds them to be predominantly filamentary. Due to their large volume, SDSS and 2dFGRS can be fruitfully employed to test predictions for MFs in a weakly nonlinear regime. An interesting application has been illustrated by Colley *et al.*[101], who use mock SDSS catalogues (inspired by ΛCDM cosmology) for the purpose.

Morphology of LSS in LCRS has been studied by Bhardwaj *et al.*[26] who detect significant filamentarity in LCRS slices compared to an equivalent random distribution. The LSS-data from recent redshift surveys such as SDSS and 2dFGRS yet await an integrated morphological analysis and a detailed comparison with theoretical predictions. Meanwhile, encouraging efforts have been recently made to constrain the scale of homogeneity of the LSS by investigating the *largest length*-scale of filaments in LCRS slices[27]. Robust 2-dimensional morphological statistic such as the Shapefinder $\mathcal{F}$ have been used for the purpose. The scale of homogeneity is found to be $\simeq 70$–$80\ h^{-1}$ Mpc. ΛCDM *mock* LCRS catalogues have been further confronted with the LCRS slices using the same techniques. A uniform, moderate large scale bias of 1.15 seems to give good agreement between the model and the LCRS slices.

1. Spergel, D. N. *et al.*, *ApJ*, 2003, **148**, 175.
2. Morphology is a combined study of sizes and shapes of structures. Connectivity, geometry and the overall shape of the structures form the subject matter of such a study.
3. Shandarin, S. F., *Sov. Astron. Lett.*, 1983, **9**, 104.
4. Shandarin, S. F. and Zeldovich, Ya. B., *Rev. Mod. Phys.*, 1989, **61**, 185.
5. Gott, J. R., Melott, A. L. and Dickinson, M., *ApJ*, 1986, **306**, 341.
6. Melott, A. L., *Phys. Rep.*, 1990, **193**, 1.
7. Barrow, J., Bhavsar, S. P. and Sonoda, D. H., *MNRAS*, 1985, **216**, 17.
8. White, S. D. M., *MNRAS*, 1979, **186**, 145.
9. van de Weygaert, R., *A&A*, 1994, **283**, 361.
10. Mecke, K. R., Buchert, T. and Wagner, H., *A&A*, 1994, **288**, 697.
11. Schmalzing, J. and Buchert, T., *ApJL*, 1997, **482**, 1.
12. Sahni, V. and Coles, P., *Phys. Rep.*, 1995, **262**, 1.
13. van de Weygaert, R., Froth across the Universe, Dynamics and Stochastic Geometry of the Cosmic Foam, 2002, arXiv: astro-ph/0206427.
14. Martinez, V. J. and Saar, E., *Statistics of Galaxy Distribution*, Chapman & Hall, London, 2002.
15. Jones, B. J. T., Martinez, V. J., Saar, E. and Trimble, T., 2004, astro-ph/0406086.
16. Shane, C. D. and Wirtanen, C. A., *AJ*, 1954, **59**, 285.
17. Hubble, E., *Proc. Natl. Acad. Sci.*, 1929, **15**, 168.
18. de Lapparent, V., Geller, M. and Huchra, J., *ApJL*, 1986, **302**, 1.
19. da Costa, L. N., Pellegrini, P. S., Davis, M., Meiksin, A., Sargent, W. L. W. and Tonry, J., *ApJS*, 1991, **75**, 935.
20. Fairall, A., *Large-Scale Structure in the Universe*, John-Wiley & Sons in association with Praxis Publishing, Chichester, 1998.
21. Santiago, B. X., Strauss, M. A., Lahav, O., Davis, M. Dressler, A. and Huchra, J., *ApJ*, 1995, **446**, 457.
22. Loveday, J., Maddox, S. J., Efstathiou, G. and Peterson, B. A., *ApJ*, 1995, **442**, 457.
23. Ratcliff, A. *et al.*, *MNRAS*, 1998, **300**, 417.
24. Shectman, S. A. *et al.*, *ApJ*, 1996, **470**, 172.
25. Colley, W. N., *ApJ*, 1997, **489**, 471.
26. Bharadwaj, S., Sahni, V., Sathyaprakash, B. S., Shandarin, S. F., and Yess, C., *ApJ*, 2000, **528**, 21.
27. Bharadwaj, S., Bhavsar, S. P. and Sheth, J. V., *ApJ*, 2004, **606**, 25.
28. Bharadwaj, S. and Pandey, B., 2004, arXiv: astro-ph/0405059.
29. Colless, M., *et al.*, astro-ph/0306581.
30. Hawkins, E. *et al.*, *MNRAS*, 2003, **346**, 78.
31. Peacock, J., 2003, astro-ph/0309240.
32. Earlier these were already estimated by refs 33 and 34, respectively by working with the 2-dimensional galaxy-counts recorded in the APM catalogue. Also see ref. 35.
33. Baugh, C. M. and Efstathiou, G., *MNRAS*, 1993, **270**, 183.
34. Baugh, C. M., *MNRAS*, 1996, **280**, 267.
35. Maddox, S., Efstathiou, G. and Sutherland, W. J., *MNRAS*, 1996, **283**, 1227.
36. Percival, W. *et al.*, *MNRAS*, 2001, **327**, 1297.
37. Tegmark, M., Hamilton, A. and Xu, Y., *MNRAS*, 2002, **335**, 887.
38. Madgwick, D., Lahav, O., Baldry, I. and the 2dFGRS team, *MNRAS*, 2002, **333**, 133.
39. Norberg, P., Baugh, C., Hawkins, E. and the 2dFGRS team, *MNRAS*, 2002, **332**, 827.
40. Lahav, O., Bridle, S., Percival, W, J. and the 2dFGRS team, *MNRAS*, 2002, **333**, 961.
41. Hoyle, F. and Vogeley, M. S., *ApJ*, 2004, **607**, 751.
42. Hoyle, F. and Vogeley, M. S., *ApJ*, 2002, **566**, 641.
43. Hoyle, F. *et al.*, *ApJ*, 2002, **580**, 663.
44. Hikage, C. *et al.*, *PASJ*, 2003, **55**, 911.
45. Basilakos, S., *MNRAS*, 2003, **344**, 602.
46. Doroshkevich, A., Tucker, D. L., Allam, S. and Way, M. J., *A&A*, 2004, **418**, 7.
47. Hikage, C. *et al.*, PASJ, 2002, **54**, 707.
48. Abazajian, K., Adelman-McCarthy, J. K., Agueros, M. A., Allam, S. S. and the SDSS Collaboration, 2004, astro-ph/0403325, Submitted to AJ.
49. Tegmark, M. and the SDSS Collaboration, *ApJ*, 2004, **606**, 702.
50. Kayo, I. *et al.*, *Publ. Astron. Soc. Jpn*, 2004, **56**, 415.

51. Tegmark, M. and the SDSS Collaboration, *Phys. Rev. D*, 2004, **69**, 103501.
52. Sheth, J. V., astro-ph/0310755, Accepted for publication in *MNRAS*, 2004.
53. Bertschinger, E., *Annu. Rev. Astron. Astrophys.*, 1998, **36**, 599.
54. Klypin, A. A., 2000, astro-ph/0005502.
55. Klypin, A. A., 2000, astro-ph/0005503.
56. Klypin, A. A., 2000, astro-ph/0005504.
57. Bond, J. R., Kofman, L. A. and Pogosyan, D., *Nature*, 1996, **380**, 603.
58. Weinberg, D. H., Hernquist, L. and Katz, N., *ApJ*, 2002, **571**, 15.
59. Weinberg, D. H., Dáve, R., Katz, N. and Hernquist, L., *ApJ*, 2004, **601**, 1.
60. Benson, A. J., Lacey, C. G., Baugh, C. M., Cole, S. and Frenk, C. S., *MNRAS*, 2002, **333**, 156.
61. Cole, S., Lacey, C. G., Baugh, C. M. and Frenk, C. S., *MNRAS*, 2000, **319**, 168.
62. Kauffmann, G., Nusser, A. and Steinmetz, M., *MNRAS*, 1997, **286**, 795.
63. Scoccimarro, R. and Sheth, R. K., *MNRAS*, 2002, **329**, 629.
64. Peacock, J. A. and Dodds, S. J., *MNRAS*, 1994, **267**, 1020.
65. Kaiser, N., *ApJL*, 1984, **284**, 9.
66. Norberg, P. Baugh, C., Hawkins, E. and the 2dFGRS team, *MNRAS*, 2001, **328**, 64.
67. Groth, E. and Peebles, P. J. E., *ApJ*, 1977, **217**, 385.
68. Klypin, A. A., Primack, J. and Holtzman, J., *ApJ*, 1996, **466**, 13.
69. Peacock, J., *MNRAS*, 1997, **284**, 885.
70. Jenkins, A. R. *et al.* (for the Virgo Constortium), *ApJ*, 1998, **499**, 20.
71. Cole, S. and Kaiser, N., *MNRAS*, 1989, **237**, 1127.
72. Mo, H. and White, S. D. M., *MNRAS*, 1996, **282**, 347.
73. Steidel, C., *Bull. Am. Astron. Soc.*, 1997, **29**, 1400.
74. Cooray, A. and Sheth, R. K., *Phys. Rep.*, 2002, **372**, 1.
75. Indeed, about 70% of galaxies appear to be in the field, i.e. along filaments and sheets connecting the clusters of galaxies.
76. The two point correlation function completely describes the properties of a Gaussian random field. Although the present distribution of both baryons and dark matter arose from an initially 'featureless' Gaussian distribution, the process of gravitational clustering gives rise to significantly non-Gaussian features (superclusters, voids, etc.) to describe which one needs to go beyond the simple two point correlation function to higher order correlation function's or, alternatively, by means of geometrical diagnostics such as the genus, percolation theory and more generally the Minkowski functionals which will be described below.
77. Dekel, A. and Lahav, O., *ApJ*, 1999, **520**, 24.
78. Blaschke, W., *Integralgeometrie. Erstes Heft, Bernd G. Teubner*, Leipzig, Berlin, 1936.
79. Minkowski, H., Volumen und Oberfläche, *Math. Ann.*, 1903, **57**, 447, in German.
80. Matsubara, T., *ApJ*, 2003, **584**, 1.
81. Schmalzing, J., Ph D thesis, Ludwig-Maxmilians-Universitat Munchen, 1999.
82. Doroshkevich, A., *Astrofizika*, 1970, **6**, 320.
83. Zeldovich, Ya. B., Einasto, J. and Shandarin, S. F., *Nature*, 1982, **300**, 407.
84. Klypin, A. A. and Shandarin, S. F., *ApJ*, 1993, **413**, 48.
85. Sahni, V., Sathyaprakash, B. S. and Shandarin, S. F., *ApJL*, 1998, **476**, 1.
86. Shandarin, S. F. and Yess, C., *ApJ*, 1998, **505**, 12.
87. Shandarin, S. F., Sheth, J. V. and Sahni, V., *MNRAS*, 2004, **353**, 162.
88. *FF* is the cumulative probability distribution function: $\mathrm{FF}(\delta T) = P(\delta > \delta_T)$.
89. de Lapparent, V., Geller, M. and Huchra, J., *ApJ*, 1991, **369**, 273.
90. Hamilton, A. J. S., Gott, J. R. and Weinberg, D. H., *ApJ*, 1986, **309**, 1.
91. Gott, J. R., Weinberg, D. H. and Melott, A. L., *ApJ*, 1987, **319**, 1.
92. Sahni, V., Sathyaprakash, B. S. and Shandarin, S. F., *ApJL*, 1998, **495**, 5.
93. Sheth, J. V., Sahni, V., Shandarin, S. F. and Sathyaprakash, B. *MNRAS*, 2003, **343**, 22.
94. Sheth, J. V., Ph D thesis, Pune University, Pune, India, 2004.
95. Schmalzing, J., Buchert, T., Melott, A. L., Sahni, V., Sathyaprakash, B. S. and Shandarin, S. F., *ApJ*, 1999, **526**, 568.
96. Cole, S., Hatton, S., Weinberg, D. H. and Frenk, C. S., *MNRAS*, 1998, **300**, 945.
97. Springel, V. *et al.* (for the Virgo Consortium), *MNRAS*, 1998, **298**, 1169.
98. Gott, J. R. *et al.*, 2003, astro-ph/0310571.
99. Sathyaprakash, B. S., Sahni, V., Shandarin, S. F. and Fisher, K. B., *ApJL*, 1998, **507**, 109.
100. It is worth mentioning here that these authors do not account for nontrivial effects of bias in their treatment.
101. Colley, W. N., Gott, J. R., Weinberg, D. H., Park, C. and Berlind, A. A., *ApJ*, 2000, **529**, 795.