



REVIEWING THE MULTIFRACTAL NATURE OF CLUSTERING OF GALAXIES IN THE UNIVERSE

Jaswant K. Yadav
 Department of Physics,
 Central University of Haryana
 Mahendergarh 123029, Haryana, INDIA

Abstract: Large aperture telescopes have been employed to determine the distribution and structures of galaxies on different length scales in the Universe. The primary information that these surveys of galaxies provide us concerns about non random nature distribution of galaxies and structures of galaxies in our neighbourhood and beyond. Nature of these structures shall be better analysed in greater details once we have more data coming in from existing surveys as well as those that are to be undertaken in future. Since we do not have direct probes of the earlier history of the universe we hope to gather necessary information from these galaxy surveys so as to describe the evolutionary processes that are responsible for the formation of these clustered structures starting from initially homogeneous distribution of gas in the universe. We review different statistical methods to study the clustering of large scale structures of our Universe with special emphasis on fractal analysis approach. We highlight with comprehensive description, the fractal versus homogeneous debate concerning the distribution of matter structures on various length scales. Various arguments for and against fractal nature of the galaxy distribution are presented. Historical development of this branch of data analysis involves employing methods in which the definition of the measure of clustering does not depend on average density of the universe. We conclude that much of the statistical analysis point to the presence of fractal nature on small scales with a transition to the homogeneous distribution of galaxies as we study their clustering on comparatively larger scales. The aim of this review is therefore to explain and to categorize the various algorithms into groups and their application in the field of astrophysical signal analysis. In order to put the discussion in right perspective this review should be helpful in differentiating bare facts from mere opinions.

Keywords: methods: statistical, computational – cosmology: theory – large scale structure of the universe; Fractal :Multifractal spectrum

I. INTRODUCTION

The standard approach to large scale matter distribution in the Universe assumes that the Universe can be modeled as a perturbed FriedmanLemaitreRobertsonWalker (FLRW) universe. The large-scale in this terminology are the scales of size more than one mega parsec (Mpc) which is about the distance traveled by light in 3.26 million years. These type of structures in the universe are believed to have been formed due to the collapse of small scale inhomogeneities present in the early Universe (Peebles 1980 [1]; Peacock 1999 [2]; Bernardeau et al 2002 [3]; Padmanabhan 2002 [4]). The growth of different structures is such that amplitudes grow proportional to the linear growth factor of the universe but positions of the density maxima do not change. Structures of different scales are correlated implying that large density structures are surrounded by large volume voids. Clusters of galaxies are believed to have formed in places where density maxima of small and medium scale overlap each other. Superclusters of galaxies are situated at a location where density maxima of clusters of galaxies overlap. Voids form in regions where large scale density perturbations combine near minima of waves. All this correlation of small scale and large scale density forms an interconnected network known as cosmic web (Einasto J. 2012 [5]). Almost all the theoretical calculations of the dynamics of universes assume that the universe is homogeneous and isotropic on large scales (Einstein 1917 [6]). This is also known as cosmological principle. Some of the indirect proofs of this principle are (i) recession velocity of galaxies being proportional to their distance (Hubble law), (ii) highly isotropic nature of cosmic microwave

background radiation as observed by COBE and (iii) the abundances of light elements which matches well with the production of elements in a region that has evolved from an initial hot and dense environment. This principle has made it easier for us to understand the growth of structures using linear perturbation theory. At a time when cosmological principle was proposed, there were hardly any observations regarding the structure and distribution of matter in the universe so that the principle could be verified. It was due to unavailability of astronomical observational tools to undertake these observations. In the last three decades, however, we have made significant progress in technology knowhow as well as scientific instrumentation so as to produce large amount of useful astronomical data in finite time. It is therefore extremely essential to test whether the observed distribution of galaxies approaches a homogeneous distribution at large scales.

There have been various attempts in last 30 years to quantify the distribution of galaxies in the night sky. The first such attempt was made by Huchra et al. (1983) [7] who observed about 2500 galaxies to an apparent magnitude of 14.5. This survey however was very sparse and hence could not be used to study the distribution in greater details. A more serious attempt in the form of Center for Astrophysics-II (CfA-II) redshift survey resulted in a far better slice of the universe which contained a large volume, and was very deep extending to about 150 Mpc. The Southern Sky Redshift Survey (SSRS) (da costa et al. 1991 [8]) was a complement to the CfA-II survey in southern sky. This survey and its followup included redshift of about 2000 and 5400 galaxies respectively. Data from the second CfA survey exhibited the presence of galaxies that were not randomly distributed but were clustered along the surface of empty "voids". The Las

Campanas Redshift Survey (Shectman et al. 1996 [9]) had a coverage over six thin parallel slices ($1.5^\circ \times 90^\circ$) with the depth about 750 Mpc. The six slices of the survey contained about 24000 galaxies. The sufficient volume coverage of the survey allowed astronomers to ponder whether our knowledge of the clustering of nearby Universe was sufficient to describe parts of the universe that are more distant from us. The findings of this survey indicated the luminosity density of galaxies to depend on galaxy number density as well as morphology of the galaxy. Attempts were also made to study second- and higher-order correlation functions of galaxy population to describe the statistical nature of galaxy clustering.

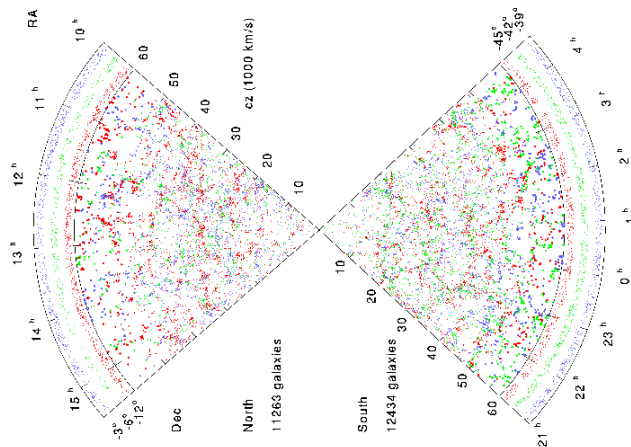


Figure 1. The representation of galaxies in the right ascension (RA) and recession velocity (cz) coordinate systems as observed in Las Campanas Redshift survey (LCRS). The color of the galaxy indicates the declination strip to which the galaxy belongs to (Figure courtesy : <http://www.astro.ucla.edu/~wright/lcrs.html>).

The 2dF Galaxy Redshift Survey (2dFGRS : Colless et al. 2001 [10]) is a major galaxy redshift survey in which the spectroscopy of various astronomical objects has been performed with the unique capabilities of the 2dF facilities built by the Anglo-Australian Observatory using the 3.9m Anglo-Australian Telescope between 1997 and April 2002. This survey obtained spectroscopic information about 245991 galaxies brighter than an apparent magnitude of 19.5 and covering a sky area of approximately 1500 square degrees. The data obtained from this survey has been used to address various issues related to galaxy formation and cosmology including the characterization of luminosity function of galaxies in visible (Norberg et al. 2002 [11]) and low energy part of electromagnetic spectrum (Cole et al 2005 [12]). The visible region luminosity function provides the mean current star-formation rate of galaxies while its low energy (near infrared) counterpart informs us about the stellar mass function of galaxies.

The Sloan Digital Sky Survey (SDSS) (York et al 2000 [13]) is one of the most well planned and path-breaking surveys undertaken in astronomy. The survey started gathering data from year 2000 onwards and has collected deep, multi-color images containing more than 930,000 galaxies and more than 120,000 quasars. Overall it has covered more than a quarter of the sky and helped us create 3-dimensional maps of the Universe. It used a dedicated 2.5-

meter telescope at Apache Point Observatory, New Mexico. The telescope has observed in both imaging and spectroscopic modes. Images were taken using a set of five filters (named *u*, *g*, *r*, *i* and *z*) with average wavelength of 355.1, 468.6, 616.5, 748.1 and 893.1 nanometers respectively. The two key technologies that enabled the SDSS, to perform to this unprecedented level are the high quality optical fibers and the digital imaging detectors known as CCDs. The SDSS helped us discover the most distant quasars, powered by supermassive black holes in the early Universe and large populations of sub-stellar objects. It also helped us make precision measurement of the luminosity distribution of quasars, large scale clustering and cosmological constraints as well as early structure with the Lyman-alpha forest.

In essence we can conclude that the galaxy redshift surveys have provided us with large number of galaxies contained in a large enough volume in the sky. The nature of clustering of these galaxies can be studied with the help of multifractal spectrum of this distribution. In section II we describe fractal analysis of galaxy distribution obtained from volume limited subsamples of galaxies. After that in Section III, we briefly describe various other methods to describe the clustering of this large scale distribution of matter in the universe. We finish this review with a set of conclusions in Section IV.

II. FRACTAL ANALYSIS OF GALAXY DISTRIBUTION

Fractals are self-similar objects which have been invoked to describe many physical phenomena which exhibit self-similarity (Mandelbrot, 1982 [14]). A multifractal can be assumed to be a generalization of the concept of a fractal. The main characteristics of a galaxy distribution that exhibits multifractal behavior is that the self-similar behavior of particle distributions may be different in different density environments. A complete statistical information about any point distribution can be obtained from the the knowledge of statistical moments of that distribution. The multifractal analysis characterizes scaling properties of moments at all levels. Mathematical quantification of a fractal or multifractal distribution is provided by the fractal dimension.

The behavior of correlation function over a range of length-scale is a power law in nature. This observation led Pietronero (1987) [15] to propose that the distribution of galaxies follows closely the fractal distribution which themselves are self-similar in nature. The analysis of a sample of galaxies by Coleman & Pietronero (1992) [16] seemed to support such a proposition.

A study by Sylos Labini et al (1998) [17] ruled the galaxy distribution to be of fractal nature with no transition to homogeneity on any length-scale. For such a distribution the mean density of universe will reduce as the volume of the galaxy survey increases. This should also manifest itself as an increase in the correlation length on progressively larger scales of observation. Further analysis of different galaxy data sets by the same group of scientists continues to hold the fractal paradigm for galaxy distribution. However the fractal paradigm does not seem to hold in the analysis of

volume-limited samples of CfA II redshift survey with increasing depth (Martínez, López-Martí & Pons-Bordería 2001 [18]). Here the authors obtained a constant value of correlation length even when the size of the analysis region is different. In a fractal analysis of the data slices obtained from European Science Observatories survey, Guzzo (1997 [19]) confirmed the large-scale homogeneity of the Universe while a similar analysis of the volume-limited samples of SSRS 2 could not distinguish between the fractal and homogeneous nature of galaxy distribution. A fractal analysis carried out on the APM-Stromlo observatory (Hatton 1999 [20]) survey was consistent with the fractal behavior of the matter distribution up to 40 Mpc length-scale. Borgani (1995) [21] via his analysis of different points distribution of galaxies claimed the validity of fractal nature on small scales while the large-scale visible matter distribution was behaving the same way as a random distribution would do on large-scales.

As we had described in Section II, the LCRS was a wide and deep survey of the Universe carried out in the last decade of 20th century. A fractal analysis of that survey by Amendola & Palladino (1999) [22] found a self-similar behavior on scales less than ~30 Mpc but the analysis was inconclusive whether the distribution actually had a transition to homogeneity. In a seminal work on the similar data set Bharadwaj, Gupta & Seshadri (1999) [23] showed via their multifractal analysis that the universe exhibits homogeneity on the scales more than 80 Mpc in size all the way to 200 Mpc in size which was the size of the particular survey. A different group (Kurokawa, Morikawa & Mouri 2001 [24]) performing a similar analysis showed that homogeneity transition scales is about 30 Mpc while Best (2000) [25] failed to report a departure from fractality even on the largest scale analysed.

Pan & Coles (2000) [26] carried out a careful analysis of Point Source Catalogue by taking into account the corrections arising due to irregular survey boundaries and showed the transition to homogeneous nature occurring on length-scale around 30 Mpc. A fractal analysis of SDSS Early Data Release by Baryshev & Bukhmastova (2004) [27] found the continuity of the fractal behavior on length-scale all the way up to 200 Mpc while an analysis of Luminous Red Galaxies (Hogg et al. 2005 [28]) seemed to suggest the distribution to be homogeneous at around 70 Mpc.

Yadav et al (2005) [29] carried out a multifractal analysis of the data obtained from SDSS Data Release-I. They used the concept of Minkowski-Bouligand dimension to determine the fractal nature of galaxy distribution. In their analysis the probability of finding a galaxy within a circle of radius r centred on another galaxy can be given as

$$C_2(r) = \frac{1}{NM} \sum n(r)$$

This definition can be generalised to obtain the general moment of a statistical distribution as

$$C_q(r) = \frac{1}{MN} \sum n(r)^{q-1}$$

Here N is the total number of galaxies and M is the galaxies on which the spheres of radius r have been kept.

$n(r)$ denotes the number of neighbors that a galaxy can have within a sphere of radius r . This generalised definition of moment can be used to define the multifractal spectrum of dimension known as Minkowski-Bouligand Dimension as

$$D_q(r) = \frac{1}{q-1} \frac{d \log C_q(r)}{d \log(r)}$$

In fact $D_q(r)$ quantifies the scaling behavior of different moments of the distribution. One of the interesting consequence of using $D_q(r)$ is that we are able to explain the clustering of galaxies in the cluster regions (for large positive value of q) as well as voids regions (for large negative value of q) at the same time. The results obtained in Yadav et al. (2005) [29] are reviewed in Figure 2 and 3 for some fixed value of q . A snippet of the code to produce $C_q(r)$ and $D_q(r)$ spectrum is given in supplementary material of this review. For further information the reader is referred to the original publication.

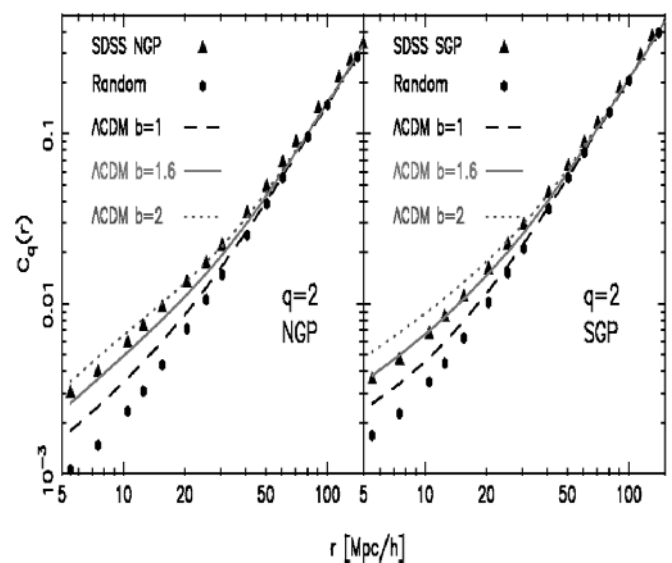


Figure 2 : This shows $C_q(r)$ at $q = 2$ for the actual data, the random data and the simulated slices. The 1σ error bars are not shown in the figure as they are too small to be seen in the logarithmic scale used here. (Figure from Yadav et al 2005)

Building upon the work of Yadav et al. (2005) [29], Bagla et al (2008) [30] obtained an expression for fractal dimension of a general point distribution and used it to find the scale of homogeneity in that distribution. The specialty of that method is that it is able to take into account the small scale clustering as well as finite number effect that are present in any redshift survey. Sarkar et al (2009) [31] extended the same analysis to 3 dimensional distribution obtained from SDSS data release-VI at arrived at similar conclusions. Yadav et al (2010) [32] used the general expression for multifractal dimension to arrive at the conclusion that an ideal redshift survey in which the size of the survey is sufficiently large should have a scale of homogeneity in excess of 200 Mpc. Scrimgeour et al (2012) [33] seemed to favor homogeneity of the distribution across a range of redshift with the help of galaxy distribution from WiggleZ redshift survey. It is to be noted that this scale is much less than the Hubble scale which is the scale of large scale

homogeneity in the cosmological principle. More recently the scale of homogeneity has been checked for multiple tracers of large-scale structures viz. main galaxies, Luminous Red galaxies and quasars sample from SDSS (Sarkar et al 2016 [34]). All these tracers point to a distribution which shows fractal behavior on smaller scales with a transition to homogeneity happening at scales around 100 Mpc.

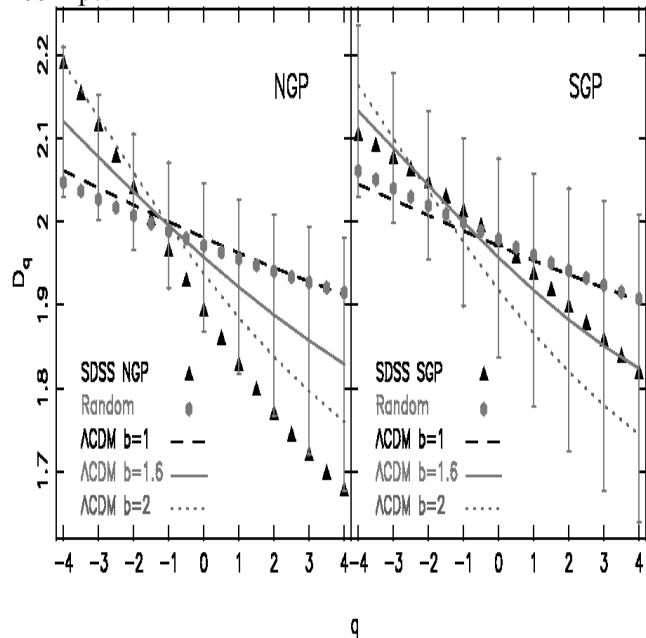


Figure 3: This shows the spectrum of generalized dimensions D_q as a function of q for the actual data from redshift survey of galaxies, the random data and the simulated slices on length-scales of 60–70 Mpc to 150 Mpc. The error bars shown are for CDM model with bias = 1.6. (Figure from Yadav et al. 2005)

III. OTHER STATISTICAL METHODS

There are many other statistical method to nature of clustering of galaxies. Two point Correlation function is one such method. It is a measure of the excess probability for two galaxies to remain at certain separation in comparison to random sample of galaxies. The only disadvantage of this method is that it assumes homogeneity of the galaxy distribution on the scales same as the size of the surveyed region. This is because it uses average matter density calculated from the galaxy survey itself. Also two kinds of galaxy distributions which are quite different from one another can have the same two-point correlation function. To overcome this challenge N-Point correlation function is studied to quantify clustering of the distribution. Another approach to study the clustering is called minimum spanning tree (MST) or minimum weight spanning tree. In this method we choose a subset of the edges of a connected graph that connects all the vertices together, without any cycles and with the minimum possible total edge weight (Sutherland & Efstathiou, 1991 [35]).

Nearest Neighbor Interaction was proposed by Badii & Politi, (1984) [36] to study clustering of galaxy distribution in which Hausdorff dimension of the galaxy distribution is used as an indicator of clustering in the distribution. One more approach that is equivalent to computing higher order

correlation function is the information theory approach called Shanon Entropy (Pandey 2013) [37].

IV. CONCLUSIONS

We have presented a review of multifractal techniques to study the distribution of large scale structures in the Universe. We have seen that there is a great debate on the fractal versus homogeneous nature of the matter distribution. This is essential because the outcome of this might change the course of approach to modern cosmological study. The standard homogeneous model of cosmology is successful, has got great predictive power and has held robustly against all the observational data including against current tests of homogeneity. Many studies point to the transition to homogeneity well within the existing redshift surveys while a few analysis continue to argue for fractal nature of the distribution. The upcoming survey with larger coverage in galaxy numbers as well as volume should be helpful in settling the debate one way or the other.

Different outcomes from galaxy surveys may also result from different ways in which fractal dimension can be defined for a points distribution. It is therefore always necessary to test the reliability of different fractal dimension estimators on a point distribution whose statistical properties are known a priori.

V. REFERENCES

1. Peebles, P. J. E. 1980, *The large-scale structure of the universe*, Princeton, N.J., Princeton University Press
2. Peacock, J. A. 1999, *Cosmological Physics*, by John A. Peacock, pp. 704. ISBN 052141072X. Cambridge, UK: Cambridge University Press, January 1999
3. Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. "Large-scale structure of the Universe and cosmological perturbation theory", 2002, Phys Repts, 367, 1
4. Padmanabhan, T. 2002, *Theoretical Astrophysics*, by T. Padmanabhan, Volume III: Galaxies and Cosmology, pp. 638. ISBN 0521562422. Cambridge, UK: Cambridge University Press
5. Einasto, J., "Large scale structure of the universe – A powerful probe for fundamental physics" 2012, International Journal of Modern Physics Conference Series, 12, 100
6. Einstein, A., "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie" 1917, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 142-152
7. Huchra, J., Davis, M., Latham, D., & Tonry, J. "A survey of galaxy redshifts. IV - The data" 1983, Astrophysical Journal Supplement Series, 52, 89
8. da Costa, L. N., Pellegrini, P. S., Davis, M., Meiksin, A., Sargent, W. L. W., & Tonry, J. L. "Southern Sky Redshift Survey - The catalog", 1991, Astrophysical Journal Supplement Series, 75, 935
9. Shectman, S. A., Landy, S. D., Oemler, A., Tucker, D. L., Lin, H., Kirshner, R. P., & Schechter, P. L. "The Las Campanas Redshift Survey", 1996, Astrophysical Journal, 470, 172
10. Colless, M., et al. "The 2dF Galaxy Redshift Survey: spectra and redshifts", 2001, Monthly Notices of Royal Astronomical Society, 328, 1039

11. Norberg P. et al., “*The 2dF Galaxy Redshift Survey: luminosity dependence of galaxy clustering*”, 2001, Monthly Notices of Royal Astronomical Society, 328, 64
12. Cole, S., et al. “*The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications*”, 2005, Monthly Notices of Royal Astronomical Society, 362, 505
13. York, D. G., et al. “*The Sloan Digital Sky Survey: Technical Summary*”, 2000, The Astronomical Journal, 120, 1579
14. Mandelbrot, B. B. 1982, *The Fractal Geometry of Nature*, San Francisco: Freeman, 1982
15. Pietronero, L. “*The fractal structure of the universe: Correlations of galaxies and clusters and the average mass density*”, 1987, Physica A, 144, 257
16. Coleman P. H., Pietronero L., “*The fractal structure of the universe*”, 1992, Phys Repts, 213, 311
17. Sylos Labini, F., Montuori, M., Pietronero, L. “*Scale-invariance of galaxy clustering*”, 1998, Phys Repts, 293, 61
18. Martínez V. J., López-Martí B., Pons-Bordería M., “*Does the Galaxy Correlation Length Increase with the Sample Depth?*”, 2001, Astrophysical Journal Letters, 554, L5
19. Guzzo L., “*Is the universe homogeneous? (On large scales)*”, 1997, New Astronomy, 2, 517
20. Hatton S., “*Approaching a homogeneous galaxy distribution : results from Stromlo-APM redshift survey*”, 1999, Monthly Notices of Royal Astronomical Society, 310, 1128
21. Borgani S., “*Scaling in the Universe*”, 1995, Phys Repts, 251, 1
22. Amendola L., Palladino E., “*The Scale of Homogeneity in the Las Campanas Redshift Survey*”, 1999, Astrophysical Journal Letters, 514, L1
23. Bharadwaj S., Gupta A. K., Seshadri T. R., 1999, “*Nature of clustering in the Las Campanas redshift survey*”, Monthly Notices of Royal Astronomical Society, 351, 405
24. Kurokawa T., Morikawa M., Mouri H., “*Scaling analysis of galaxy distribution in the Las Campanas Redshift Survey data*”, 2001, Astronomy & Astrophysics, 370, 358
25. Best J. S., “*An Examination of the Large-Scale Clustering of the Las Campanas Redshift Survey*”, 2000, Astrophysical Journal, 541, 519
26. Pan J., Coles P., “*Large-scale cosmic homogeneity from a multifractal analysis of the PSCz catalogue*”, 2000, Monthly Notices of Royal Astronomical Society, 318, L51
27. Baryshev Y. V., Bukhmastova Y. L., “*The Method of a Two-Point Conditional Column Density for Estimating the Fractal Dimension of the Distribution of Galaxies*”, 2004, Astronomy Letters, 30, 444
28. Hogg D. W., Eisenstein D. J., Blanton M. R., Bahcall N. A., Brinkmann J., Gunn J. E., Schneider D. P., “*Cosmic Homogeneity Demonstrated with Luminous Red Galaxies*”, 2005, Astrophysical Journal, 624, 54
29. Yadav, J., Bharadwaj, S., Pandey, B., & Seshadri, T. R. “*Testing homogeneity on large scales in the Sloan Digital Sky Survey Data Release One*”, 2005, Monthly Notices of Royal Astronomical Society, 364, 601
30. Bagla J. S., Yadav J., Seshadri T. R., “*Fractal dimensions of a weakly clustered distribution and the scale of homogeneity*”, 2008, Monthly Notices of Royal Astronomical Society, 390, 829
31. Sarkar P., Yadav J., Pandey B., Bharadwaj S., “*The scale of homogeneity of the galaxy distribution in SDSS DR6*”, 2009, Monthly Notices of Royal Astronomical Society, 399, L128
32. Yadav J. K., Bagla J. S., Khandai N., “*Fractal dimension as a measure of the scale of homogeneity*”, 2010, Monthly Notices of Royal Astronomical Society, 405, 2009
33. Scrimgeour, M. I., Davis, T., Blake, C., et al. “*The WiggleZ Dark Energy Survey: the transition to large-scale cosmic homogeneity*”, 2012, Monthly Notices of Royal Astronomical Society, 425, 116
34. Sarkar, P., Majumdar, S., Pandey, B., Kedia, A., & Sarkar, S. “*The many scales to cosmic homogeneity: Use of multiple tracers from the SDSS*”, 2016, arXiv:1611.07915
35. Sutherland W., Efstathiou G., “*Correlation functions of rich clusters of galaxies*”, 1991, Monthly Notices of Royal Astronomical Society, 248, 159
36. Badii R., Politi A., 1984, “*Hausdorff dimension and uniformity factor of strange attractors*”, Physical Review Letters, 52, 1661
37. Pandey, B. “*A method for testing the cosmic homogeneity with Shannon entropy*”, 2013, Monthly Notices of Royal Astronomical Society, 430, 3376