PLOTTING AND ANALYZING DATA RELEASE 1 GALAXIES FROM THE SLOAN DIGITAL SKY SURVEY.

Jorge A. Munoz
Physics Undergraduate, Physics Department, University of Texas at El Paso, 500 West University Avenue, El Paso, TX 79968

Keith Teesdale
Computer Science Undergraduate, Mathematics and Computer Science Department, South Carolina State University, 300 College Street NE, Orangeburg, SC 29117

Shavon Fleet
Physics Undergraduate, Physics and Engineering Physics Department, Morgan State University, 1700 East Cold Spring Lane, Baltimore, MD 21251

Mentor:
Dr. Daniel M. Smith
Associate Professor of Physics, Physical Sciences Department, South Carolina State University, 300 College Street NE, Orangeburg, SC 29117

ABSTRACT

We present measurements of galaxy clustering with information from the Sloan Digital Sky Survey. Our sample consists of 58,715 galaxies with redshifts 5,700 km $s^{-1} \leq cz \leq$ 39,000 km s^{-1}, covering approximately 2,099 square degrees. For the full sample the two-point correlation function has r_0 of 4.4 h^{-1} Mpc and γ of 1.3. When we divided the sample by color, the early-type galaxies showed a stronger and steeper correlation function than late-type galaxies, meaning that they are more intensely clustered. We created 2 and 3-dimensionals plots for the full sample and sub-samples where the clustering of galaxies and the scarceness of early-type galaxies relative to late-type galaxies can be appreciated.

1. INTRODUCTION

1.1 Sloan Digital Sky Survey

The Sloan Digital Sky Survey (SDSS) will map in detail one-quarter of the entire sky, determining the positions and absolute brightness of more than 100 million celestial objects. The SDSS telescope is located at Apache Point Observatory, Sunspot, New Mexico and is operated by the Astrophysical Research Consortium (ARC). The telescope uses a drift-scanning, mosaic CCD camera and images the sky in five different color bands by taking images through five color Filters, between 3,000 and 10,000 Ångstroms (u, g, r, i, z).

One primary goal of the Sloan Digital Sky Survey (SDSS) is to image 10,000 square degrees of the North Galactic Cap. As of July 2003 the SDSS has imaged around 2,099 square degrees of sky and taken spectra of approximately 186,245 objects including 134,00 galaxies, 17,700 Quasars (redshift < 2.3), 980 Quasars (redshift > 2.3), 17,000 stars and 4500 M stars. We are using part of this data by calculating the two-point correlation function to determine the characteristics of galaxy clustering.

The validated data is being released at annual intervals. Each release includes sufficient information to allow statistical analysis (Abazajian, et al., 2003). We used DR1, the first major data release that provides images, imaging catalogs, spectra and redshifts for download. More information on DR1 can be found in the web site http://www.sdss.org/dr1.

1.2 Overview

Seeing and mapping the structure of the Universe is difficult because of our location within the Milky Way and other phenomena such as interstellar dust. Through the use of the Sloan Digital Sky (SDSS), we were able to graphically convey an idea of the Structure of the Universe. The SDSS has only collected data from the Northern Galactic Cap. With the SDSS, astronomers were able to see, with the aid of data without precedents in quality and quantity, that galaxies are located generally in clusters and filaments separated by voids. The discovery of the different ways that galaxies cluster in space has led Sloan Digital Sky Survey researchers to new insights into the evolution of galaxies and matter in the universe.

In order to appreciate the distribution of galaxies, as part of the large-scale structure of the universe, we acquired DR1 by using the SDSS Query Analyzer (sdssQA). This tool, which works with Structured Query Language, lets the user download information from the Catalog Archive Server (CAS). Using the celestial coordinates of the galaxies gathered from DR1, galaxy positions were plotted using *Mathematica* software to create 2 and 3-Dimensional graphs of the structure of the Universe.

Our research concentrates on both early (elliptical, lenticular) and late (spiral, barred spiral, irregular) galaxies. To determine whether a galaxy is either early or late, we used the equation: $g-r = -(u-g) + 2.2$ from Strateva, et al. (2001), where g, r and u are apparent magnitudes. Finally,

the two-point correlation function was used to measure the patterns of clustering that the galaxies follow. Unlike researchers at S.C.S.U. last year, we did not use the SDSS eClass to separate spiral and elliptical galaxies; we used more galaxies for our research (58,715 instead of 21,707) and also used 1 Mpc-sized bin for the distances between each pair of galaxies.

1.3 Correlation Function

The redshift-space clustering is a central concern in observational cosmology, and it must be considered in every theory that attempts to explain the origin and evolution of the universe. In order to formulate a theory of galaxy clustering, a quantitative expression to explain the tendency of galaxies to form clusters is necessary. Galaxies within a cluster are closer to each other than to others outside it. Groth, et al. (1977) indicate that this tendency can be expressed in terms of correlation functions.

The two-point correlation function used in this research measures the tendency for a pair of galaxies to be closer than the average. The three-point and four-point correlation functions measures the same tendency between triplets and quadruplets of galaxies.

The distance between every two galaxies within the catalog is calculated and assigned to a bin. The number of pairs of galaxies within each bin is divided by the number of pairs that can be expected from corresponding bins in a random, uniform distribution. If the distribution were totally uniform, with no clustering at all, the ratio would be 1. The result is adjusted to zero by subtracting 1. If clusters are present, the function is greater than zero and measures the strength of clustering. For a complete analysis of galaxy clustering, the correlation function must be calculated for a large number of galaxies.

2. DATA

2.1 Description of the Sample

Equatorial coordinates and relative magnitudes for *u*, *g* and *r* passbands were obtained from the Sloan Digital Sky Survey (SDSS). To avoid the problem of evolution of the luminosity function, all the galaxies are 5,700 km $s^{-1} \leq cz \leq$ 39,000 km s^{-1}. This reduced our data from 130,475 galaxies to 74,283 (≈43%). The flux limit of the spectroscopic survey in r is ≈ 17.66; in order to avoid saturation on the spectrograph, and because of the flux limit, the apparent magnitude of our galaxies is chosen to be $14.5 < m_r < 17.6$. This reduced the data to 65,815 galaxies (≈50%). We wanted to study only luminous galaxies, so we made another cut in the red absolute magnitude, $-22 < M_r < -19$. This last cut left us with the final 58,715 galaxies (≈45%). For this cut, we used the formula for absolute magnitude

$$M = m - DM(z) - K(z), \quad (1)$$

where M is the absolute magnitude, m is the apparent magnitude, DM(z) is the distance modulus and K(z) is the K-correction. The redshift of a galaxy is not a linear measure of an object's distance at a moderate redshift, and the comoving distance depends on the cosmology assumed. Throughout this paper, we assume a Cold Dark Matter – Dark Energy model (ΛCDM) with Ω_m = 0.27, Ω_Λ = 0.73 and Ω_k = 0. For the comoving distance, we used the most accurate value at the time for the Hubble parameter, 0.71 (all values from C.L Bennett, et al., 2003). To obtain the comoving distance, D_c, we used the standard formula tabulated by Hogg (2000) and assumed a uniform Hubble flow:

$$D_c = D_H \int_0^z \frac{1}{\sqrt{\Omega_M (1+z)^3 + \Omega_k (1+z)^2 + \Omega_\Lambda}} dz \quad , \tag{2}$$

where D_H is the Hubble Distance, the speed of light c times the Hubble time and

$$D_H = 3000\, h^{-1}\ \text{Mpc}, \tag{3}$$

Because we observe the galaxies in a fixed observed wavelength band, K-correction is necessary to account for the fact that the observed band corresponds to different rest-frame bands at different redshifts, as Blanton, et al. warn. These corrections are dependent on the galaxy spectral energy distribution (SED), but we used the formula

$$K(z) = 2.5 \log_{10} (1 + z), \tag{4}$$

as Lin et al. (1996) say it is appropriate in the mean for an r-band-selected galaxy sample.

2.2 Treatment of Data

The measurements of redshift (z) and right ascension (ra) were used to generate two-dimensional polar plots in redshift space (Fig. 1,3-8). Three-dimensional plots in real space required measurements of redshift (z), right ascension (ra) and declination (dec). The graphical representation of the galaxies in real space was achieved by translating redshift (z) into the comoving distance (D_c) using the formula noted above.

Cartesian co-ordinates were generated using the comoving distance with the ra and dec angle values. We have transformed from spherical to Cartesian coordinates using the following:

$$x_c = D_c(z)\text{Sin}(90\text{-}dec)\text{Cos}(ra) \tag{5}$$

$$y_c = D_c(z)\text{Sin}(90\text{-}dec)\text{Sin}(ra) \tag{6}$$

$$z_c = D_c(z)\mathrm{Cos}(90\text{-}dec) \tag{7}$$

The three-dimensional representation of the galaxies did give us an idea of the large-scale structure of the universe. Clustering can be seen in the galaxies with agglomerations and voids. This shows that the universe is not uniformly distributed. When creating two-dimensional polar plots in redshift space, the following parameters were used:

$$-5° < \mathrm{dec} < 5°, \tag{8}$$

$$0° < \mathrm{ra} < 360°, \tag{9}$$

$$z < 0.1. \tag{10}$$

The separation of the galaxies into spiral and elliptical types (Fig. 2) was done using the color separation techniques of Strateva, et al. (2001). Their study was of the optical colors of 147,920 galaxies brighter than g = 21. The distribution of galaxies in the g-r vs. u-g color-color diagram is strongly bimodal, with an optimal color separator of u-r = 2.22. They used visual morphology and spectral classification of sub-samples of 287 and 500 galaxies respectively, to show that the two peaks correspond roughly to early (E, S0, Sa) and late (Sb, Sc, Irr) type galaxies.

Galaxies were separated in order to obtain and appreciate the difference between their redshift correlation functions. In contrast with Zehavi, et al. (2002) who separated galaxies with u-r = 1.8, we followed Strateva et al. (2001) methods. The scarceness of late type galaxies close to the Milky Way can be seen in both polar and Cartesian graphs.

A C++ program was written to calculate the distance between every pair of galaxies and assign them to bins according to their value. The length of each bin is 1 h^{-1}Mpc. Ten random

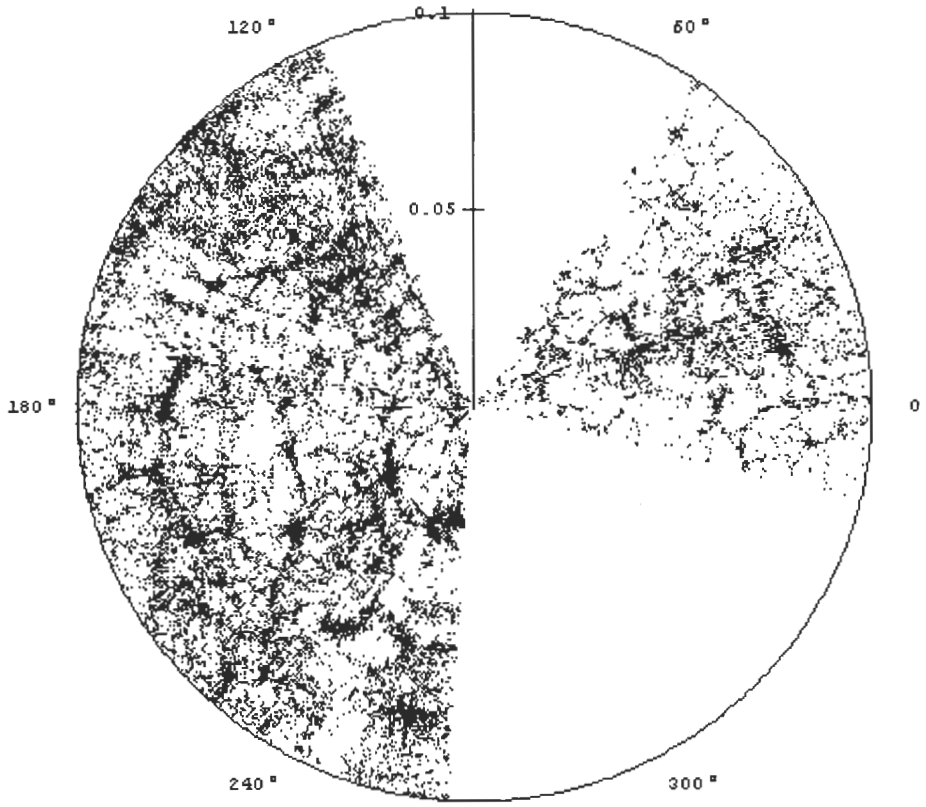

Fig 1. - Polar Plot of 35,673 Elliptical and Spiral Galaxies in Redshift Space.. -5° < dec <5°

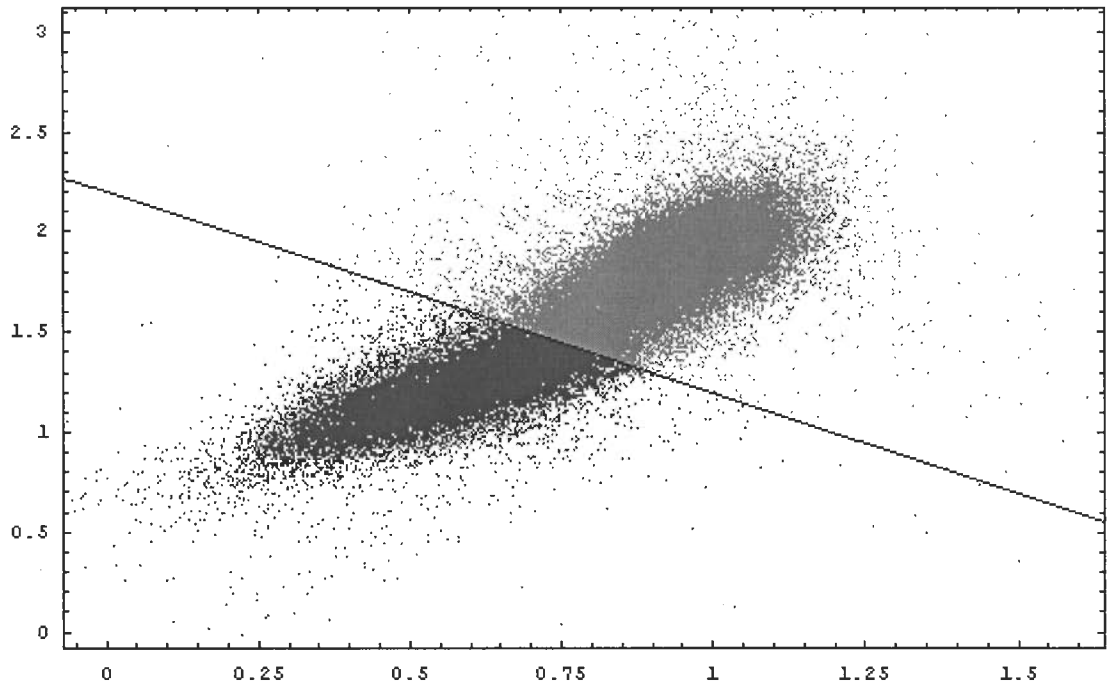

Fig. 2. – Color separation of 58715 spiral and elliptical galaxies (65:35) along Strateva et al. (2001) line of u-r line of 2.22. The x-axis represents u-g and the y-axis is g-r.

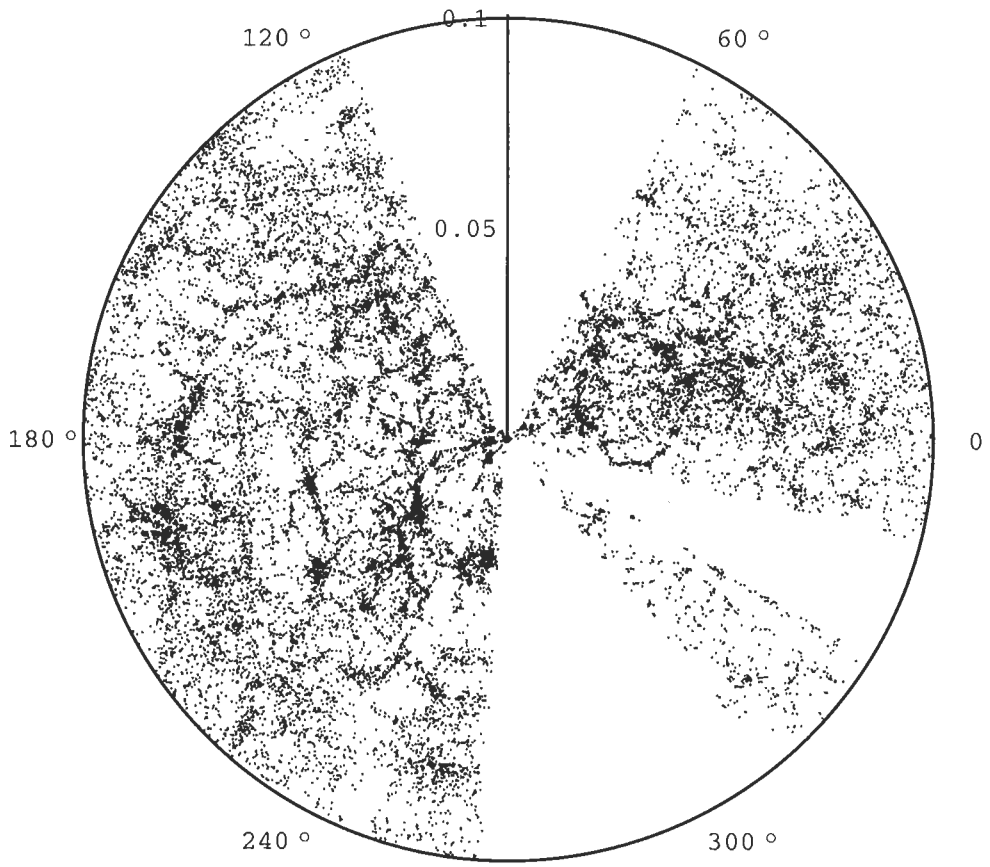

Fig. 3. – Polar plot of 24,254 spiral galaxies in redshift space, -5° < dec <5°

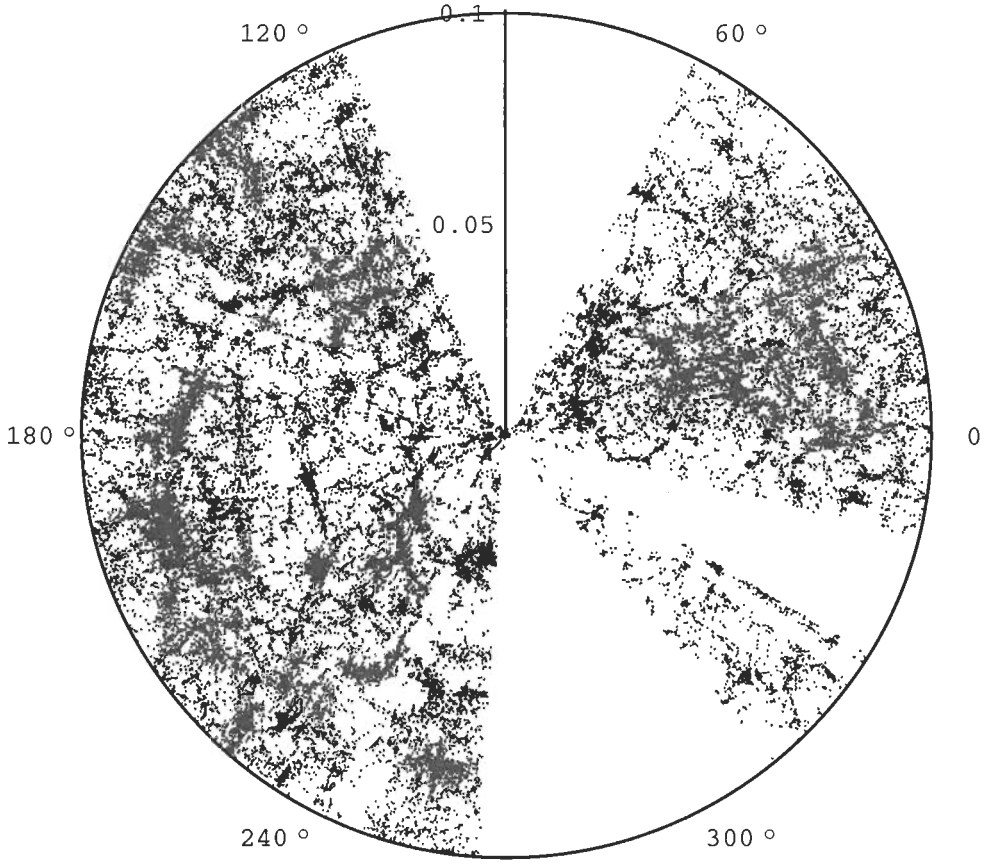

Fig. 4. – Polar plot of 31,962 elliptical galaxies in redshift space, -5° < dec <5°

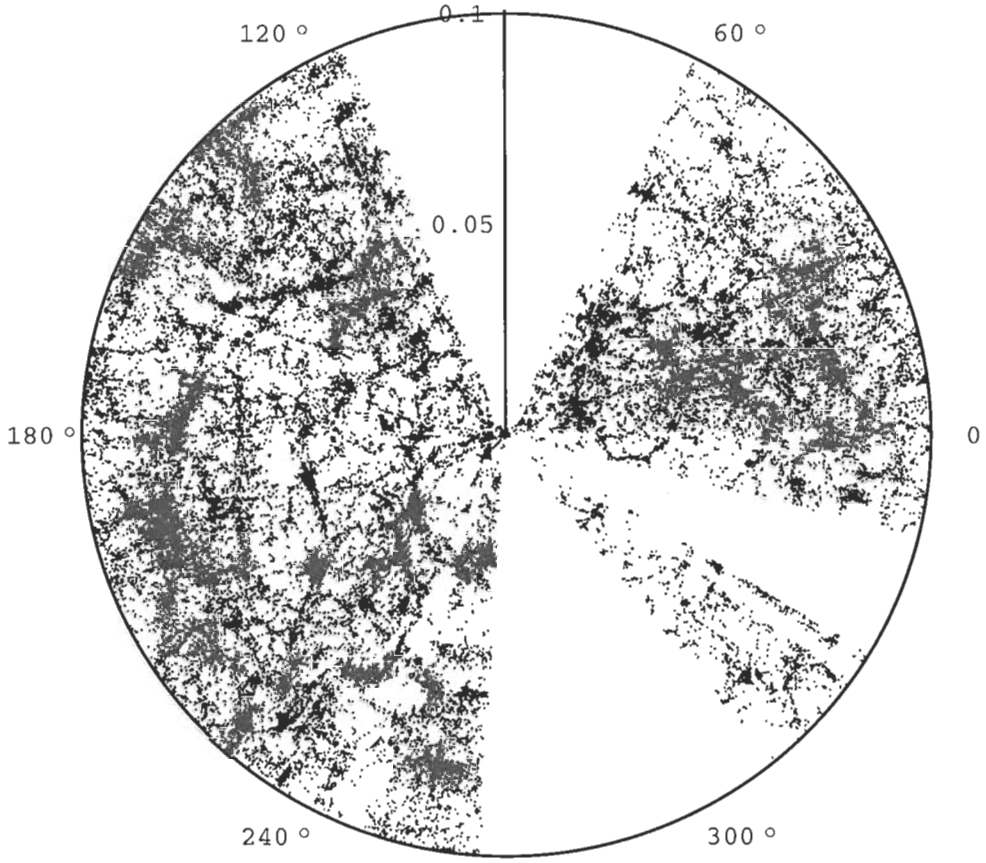

Fig. 5. – Polar plot of 56,216 galaxies in redshift space. Is important to notice the scarceness of red galaxies towards the center, where the Milky Way is located. The stronger appearance of elliptical galaxies is an artifact of layering the two images, -5° < dec <5°

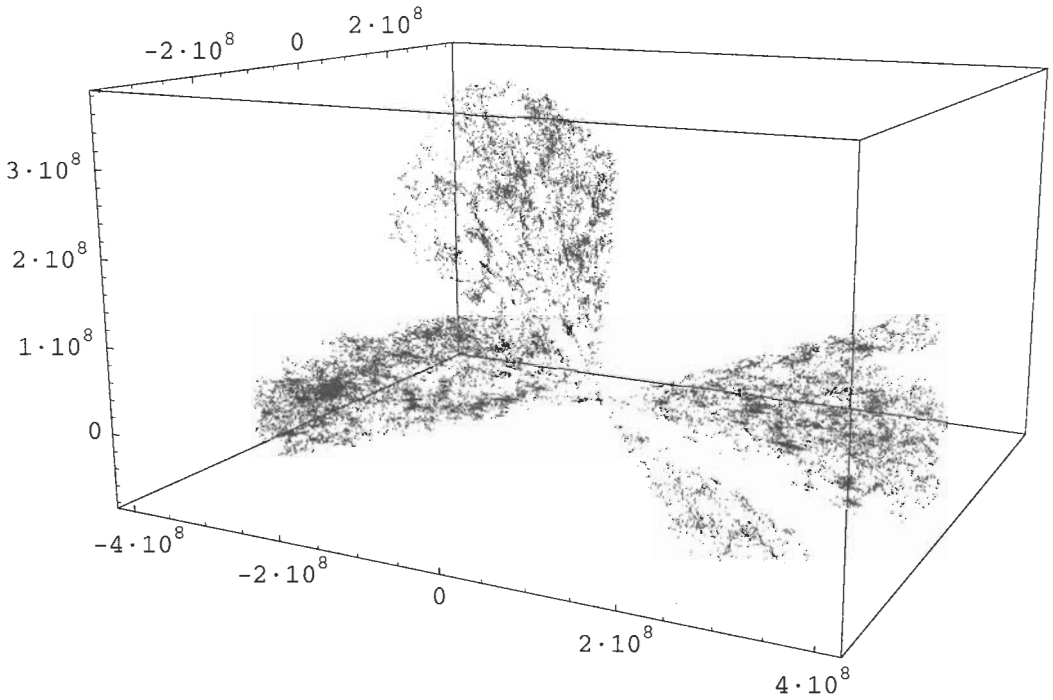

Fig. 6. – Plot of 31,962 elliptical galaxies in real space (Mpc.)

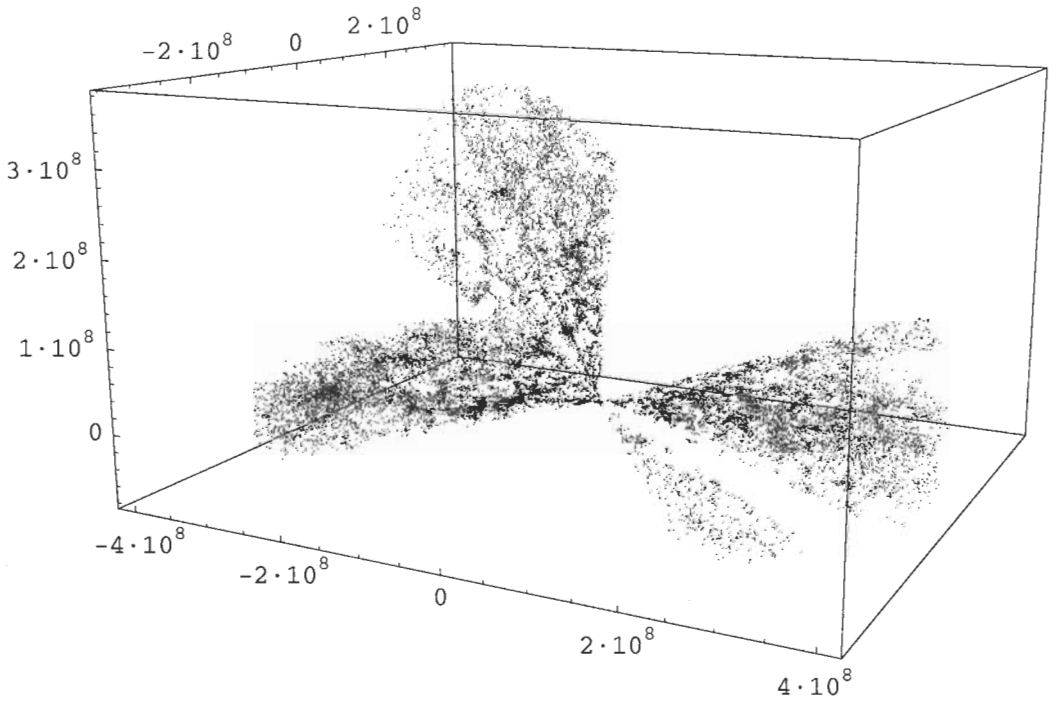

Fig. 7. – Plot of 24,254 spiral galaxies in real space (Mpc.)

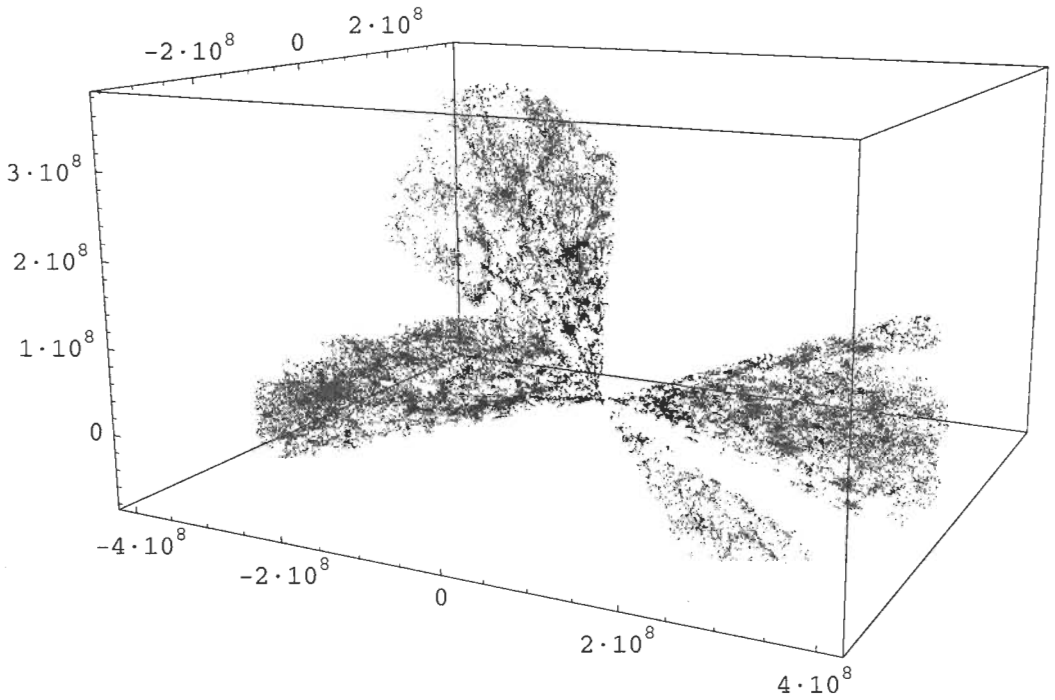

Fig. 8. – Plot of 56,216 spiral and elliptical galaxies in redshift space. Notice that the spiral (blue) galaxies are more common in the center of the graph, where the Milky Way is located. Elliptical (red) galaxies are widespread towards the edges.

catalogues were created for the main sample and sub-samples using the random number generator feature of the *Mathematica* software. This method was used to generate random z values for every ra, dec pair. When the distances were calculated and assigned to bins, they were used to calculate the 2-point spatial correlation function.

3. CORRELATION FUNCTION

3.1 Correlation formula

To account for the survey geometry, we generated random catalogs of galaxies with the same survey geometry as the real sample. We calculated the correlation function using a simple formula described in words by Groth et al. (1977),

$$\xi(r) = \left(\frac{Number\ of\ Pairs\ of\ Galaxies\ at\ r}{Number\ of\ Random\ Pairs\ of\ Galaxies\ at\ r} \right) - 1 \qquad (11)$$

where r is the comoving distance and ξ (r) tells us the strength of galaxy clustering.

We calculated the distance between every pair of galaxies in the sample and assigned it to a bin of size 1 h^{-1}Mpc. We repeated the procedure for elliptical, spiral and every random catalog.

3.2 Clustering of the Full Sample

For separations 2 h^{-1}Mpc<r<7 h^{-1}Mpc, the observed correlation function can be approximated by a power-law,

$$\xi(r) = (r/r_0)^{-\gamma} \qquad (12)$$

with $\gamma = 1.3$ and $r_0 = 4.4\ h^{-1}$ Mpc (Fig. 9). The constants γ and r_0 were determined by a fit of

$$\log \xi(r) = -\gamma(\log (r) - \log (r_0)) \qquad (13)$$

Table 1 summarizes our result for the full sample together with other important surveys. Our results should be closer to SDSS since we both are assuming a ΛCDM model (Fig. 10), in contrast with LCRS and PSCz who assume that $\Omega_m = 1$. We account for the difference in r_0 to small differences in the utilized values of Ω_m and Ω_Λ and the fact that we did not employ the radial selection function in our random catalogs, therefore, inducing some bias. Our $\xi(r)$ remains measurable non-zero out to r = 18 h^{-1} Mpc, but is no longer approximated by the power-law.

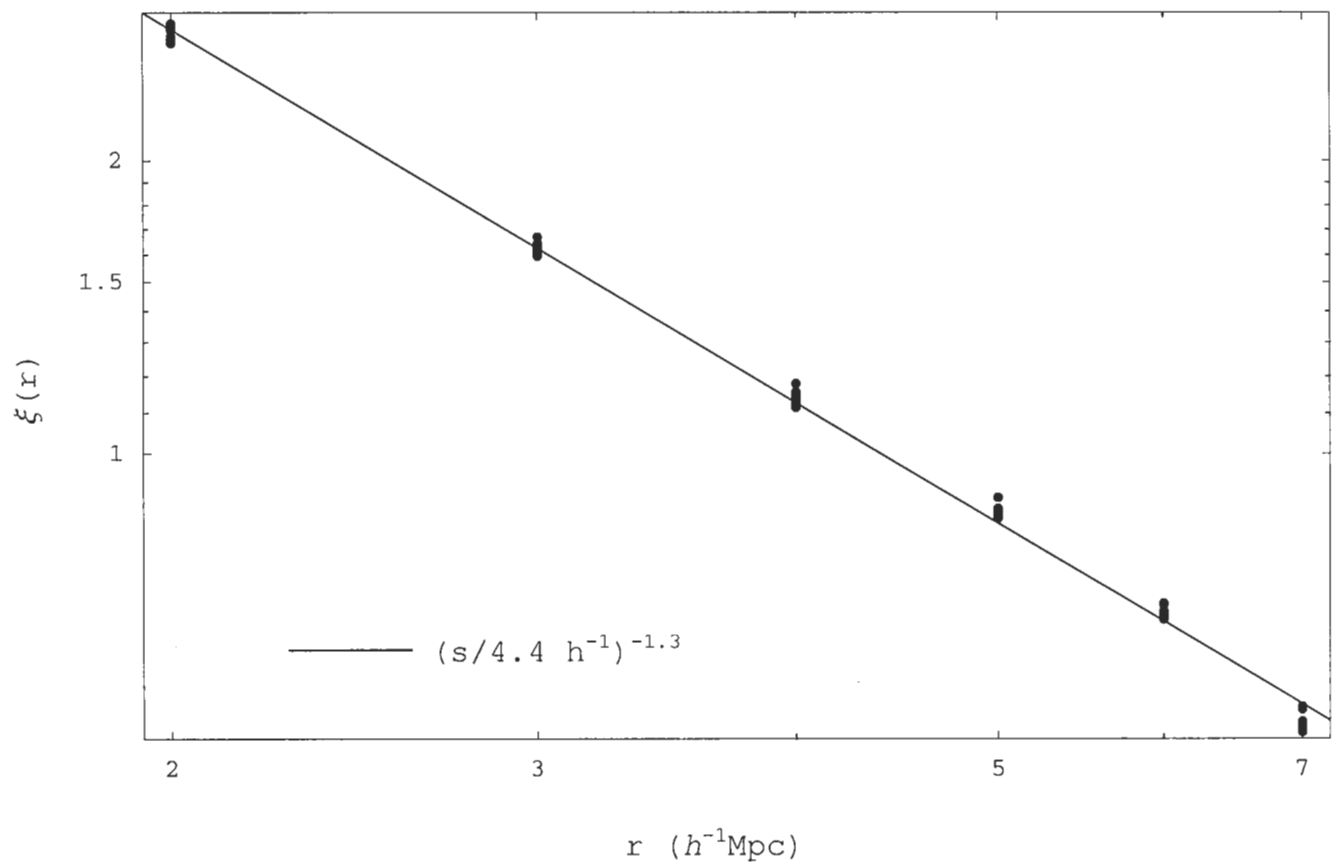

Fig. 9. – Two-point correlation function ξ(r). The lengths of the bars represent the error bounds. The solid line represents the best fit for the range 2 h^{-1}Mpc<r<7 h^{-1}Mpc.

Survey	N_{gal}	r_0	γ
SCSU[a]	58,715	4.4 ± 0.1	1.3 ± 0.1
SDSS[b]	29,300	~ 8.0	~ 1.2
LCRS[c]	26,400	6.3 ± 0.3	1.52 ± 0.03
PCSz[d]	15,400	5.0	1.2

Table 1. – Clustering results of different Galaxy Redshift Surveys. r_0 is in units of h^{-1} Mpc.

[a] Assuming a ΛCDM model with densities $\Omega_m = 0.27$ and $\Omega_\Lambda = 0.73$.
[b] Zehavi et al. 2002. ΛCDM model with $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$.
[c] Tucker et al. 1997; Jing et al. 1998 EdS model.
[d] Jing et al. 2001 EdS cosmology.

Catalog	N_{gal}	r_0	γ
Full Sample	58,715	4.4 ± 0.1	1.3 ± 0.1
Elliptical Galaxies	38,576	5.0 ± 0.2	1.3 ± 0.1
Spiral Galaxies	20,139	4.2 ± 0.1	1.2 ± 0.1

Table 2. – Best fits to the power law for different galaxy types. Notice the noticeable lower values for late-type galaxies.

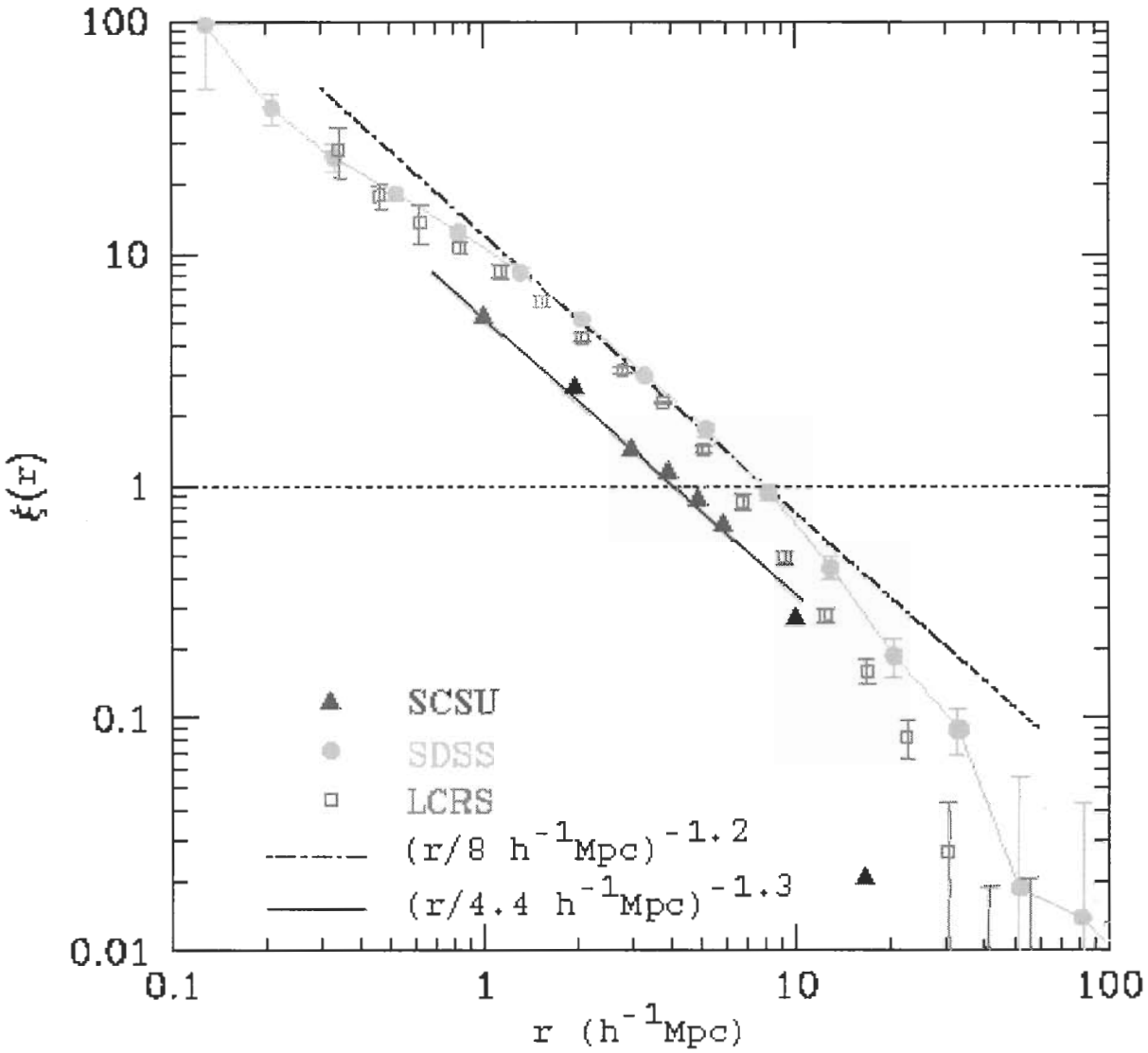

Fig. 10. – The triangles represent our measurements of the correlation function plotted against the SDSS and LCRS groups (circles and squares). The solid line represents our fit for the correlation function and the dotted line represents SDSS's. The circles' and squares' bars represent the error bounds. Notice the different amplitude but the similar general shape between all the plots. This plot was taken from Zehavi, et al. (2002). We superimposed our data into their graph.

3.3 Clustering in Spiral and Elliptical Galaxies.

We applied the same method for elliptical and spiral galaxies in the sample, separated by the techniques described above. As expected, γ and r_0 values are higher for elliptical and lower for spiral galaxies (Table 2), but the general shape of the correlation function is very similar (Fig. 11). The correlation length for sub-samples is the same as for the full sample, $2\ h^{-1}$Mpc $< r < 7\ h^{-1}$Mpc. This is in good agreement with the generally accepted theory that early-type galaxies preferentially inhabit high-density regions, as proposed by Dressler (1980).

4. CONCLUSION

We have presented measurements of galaxy clustering from SDSS DR1, based on a sample of 58715 galaxies. Since this sample covers a limited number of galaxies, our analysis has focused on galaxy clusters near the Milky Way. The sample used for this analysis is limited in radial velocity by 5,700 km $s^{-1} \leq cz \leq$ 39,000 km s^{-1}, by the flux-limit at bright and faint magnitudes at $14.5 < m_r < 17.6$ and by the absolute magnitude at $-22 < M_r < -19$. The large-scale structure of the universe can be noticed in the 2 and 3 Dimensional graphs that we created, in particular that strong clustering and the scarceness of late-type galaxies around the Milky Way.

For the full sample and sub-samples of early and late type galaxies, we have measured the redshift-space two-point correlation function. Approximating this function by a power law, $\xi(r) = (r/r_0)^{-\gamma}$, yields a correlation length r_0 of 4.4 h^{-1}Mpc and a slope γ of 1.3, but this representation is not accurate over very large scales.

Our results for the full galaxy sample are in fairly good agreement with those obtained by other research groups, in particular for the shape of the function. Restricting the analysis to the sub-sets of early and late type galaxies, we obtained similar correlation functions that are in good agreement with the hypothesis that early-type galaxies tend to inhabit high-density regions more than late-type galaxies.

The measurements of the correlation functions and other characterizations of galaxy clustering will improve in precision and detail as the SDSS progresses, yielding new information with which to understand galaxy formation.

We acknowledge NASA for support and funding through MU-SPIN Cooperative Agreement NCC 5-534 awarded to South Carolina State University.

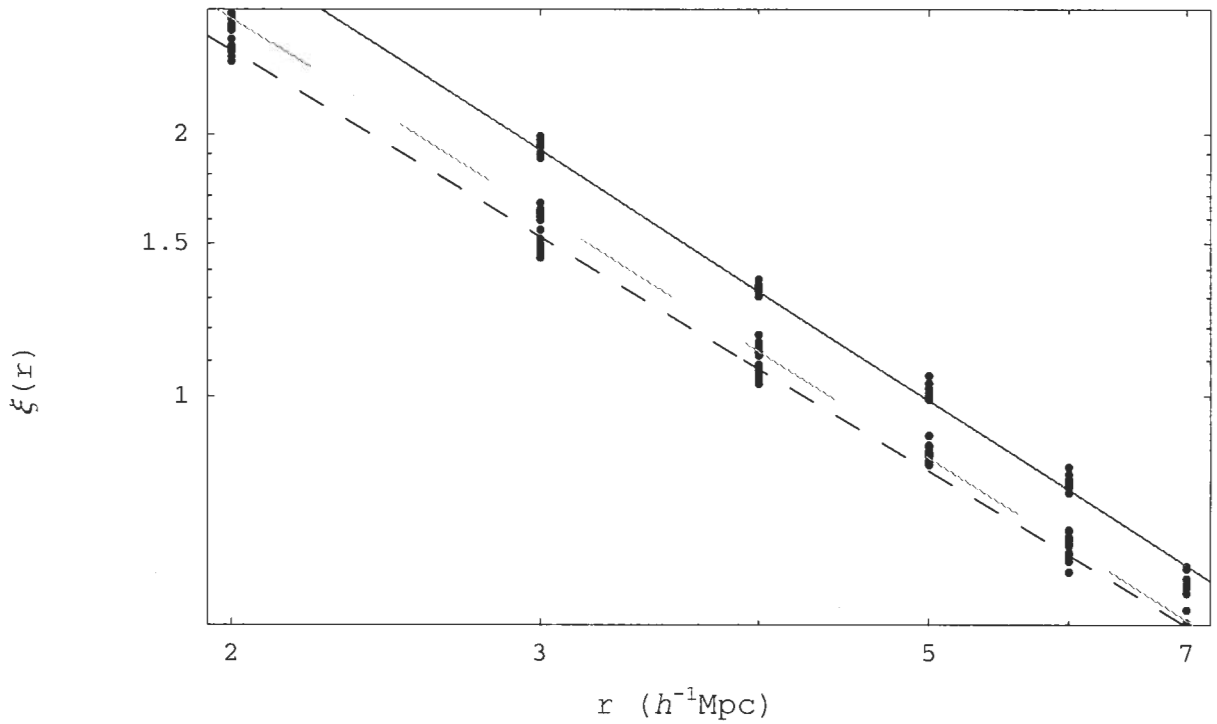

Fig. 11. – The two-point correlation function best fit for late-type galaxies (dashed line) represented by $(r/4.2\ h^{-1}\mathrm{Mpc})^{-1.2}$, early-type galaxies (solid) represented by $(r/5.0\ h^{-1}\mathrm{Mpc})^{-1.3}$ and the full sample (dot-dash) represented by $(r/4.4\ h^{-1}\mathrm{Mpc})^{-1.3}$. Since the early-type sample has higher values for r_0 and γ, the clustering is stronger among them.

REFERENCES

Abazajian, K., et al. 2003, astro-ph/0305492

Bennett, C.L., et al. 2003, astro-ph/0302207

Blanton, M.R., et al. 2001, astro-ph/0012085

Dressler, A. 1980, ApJ, 236, 351

Groth, E.J., Peebles, P.J.E., Seldner, M., Soneira, R.M. 1977, SciAm 237, p. 76-78, 84, 87-90

Hogg, D.W. 2000, astro-ph/9905116

Jing, Y.P., Börner, G., & Suto, Y. 2001, astro-ph/0104023

Jing, Y.P., Mo, H.J., & Börner, G. 1998, ApJ, 494, 1

Lin, H., et al. 1996, astro-ph/9602064

Strateva, I., et al. 2001, astro-ph/0107201

Tucker, D.L., et al. 1997, MNRAS, 285, L5

Zehavi, I., et al. 2002, astro-ph/0106476