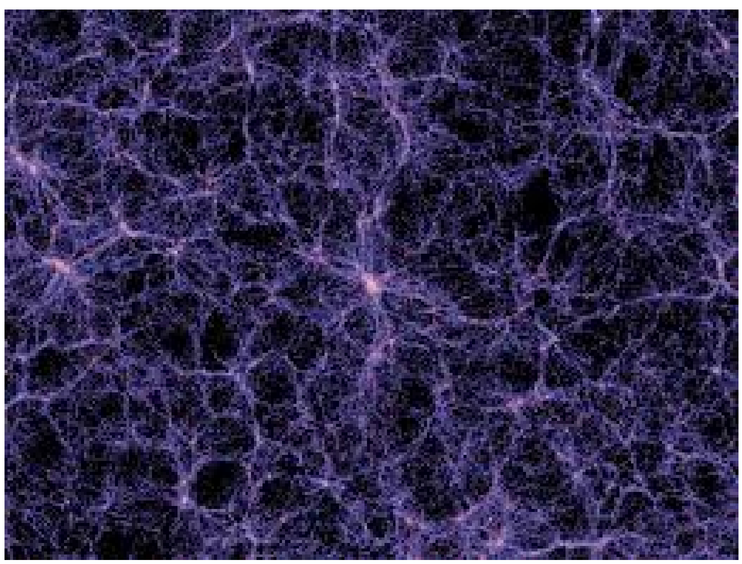
Investigating the Large-Scale Structure of the Universe

Jess Kroboth

Jeff Nechleba

December 9, 2015



Introduction

The large-scale structure of the universe is comprised of a complex topology of galaxy clusters and superclusters along with walls and filaments dispersed through massive voids. This "cosmic web" can be directly observed with data from major sky surveys such as the Sloan Digital Sky Survey (SDSS).

In this project, the large-scale structure of the universe was observed. First, data from data release 11 (DR11) of SDSS were collected and plotted. The section of the universe investigated was a redshift (z) range from 0.0 to 0.5, right ascension (RA) range from $0°$ to $360°$, and declination range from $-10°$ to $+10°$. After the large-scale structure of the universe was observed, the galaxy clusterings were mathematically characterized with a two-point correlation function. The result of the two-point correlation function was consistent with previous findings. Using the two-point correlation function, the power spectrum was found to qualitatively agree with previous findings.

Code Description

SDSS DATA

First, data were retrieved from CasJobs, a common way to query SDSS data. Four columns of interest were taken from DR11 using SQL, which are: redshift (z), right ascension (ra), declination (dec), and unique object ID (bestObjID). After specifying the desired ranges for z, RA, and Dec, data for approximately 450,000 galaxies were pulled from the database. The z, RA, and Dec columns retrieved were first extracted into individual arrays. The redshift was converted into distance measured in Mly using a known conversion rate. This made it possible to transform the RA values of each galaxy into corresponding x- and y-positions using the sine and cosine of the RA along with the found distance in Mly.

These positions were plotted to show the entire data set, which covered a span of about 12,000 Mly. However, due to the fact that the median redshift of SDSS galaxies is 0.1, the galaxies were noticeably centralized in this plot. In order to get a better feel for the concentration of the galaxies, the data were examined on a scale of 4,000 Mly. At this level the structures of galaxy filaments and clusters were much more apparent. A test of reliability of our plotted data was a comparison to SDSS published data at similar ranges. This shows the alignment of a large galaxy filament known as the Sloan Great Wall, which is discussed further in the Test Cases section.

Two-Point Correlation Function

The galaxy clustering in the large-scale structure of the universe can be characterized quantitatively. A measure for this galaxy clustering is given by the two-point correlation function. Physically, the two-point correlation function gives the relative probability of a pair of galaxies to be separated by a distance r. For instance, imagine there does not exist any gravity in the universe. A good assumption would be that the galaxies are homogeneously distributed throughout space. Therefore, given N galaxies arranged in a volume V of space, the probability of selecting two galaxies separated by a distance r would be proportional to $\frac{N^2}{V^2}$. Now imagine a universe with gravity. Because gravity is an attractive force, the clustering of galaxies would increase. This, in turn, would increase the probability of finding a pair of galaxies separated by a distance r. This relative increase in the probability of finding a pair of galaxies is just the quantity called the two-point correlation function. Mathematically it is formulated as

$$\xi(r) = \frac{2N_r}{N_d - 1}\frac{dd(r)}{dr(r)} - 1$$

where N_r is the number of galaxies in the dataset, N_d is the number of galaxies in a randomly distributed data set (with gravity turned off), $dd(r)$ is the number of galaxies separated by a distance r, and $dr(r)$ is the number of pairs of galaxies from the data and the randomly distributed data set separated by a distance r.

The code of the two-point correlation function follows closely the physical picture presented above. A function was first written to calculate the x and y coordinates of a set of galaxies constrained by redshift ($0.0 < z < 0.01$), right ascension ($0.0\,\mathrm{deg} < \mathrm{ra} < 60\,\mathrm{deg}$) and declination ($-3\,\mathrm{deg} < \mathrm{dec} < 3\,\mathrm{deg}$) obtained from the Sloan Digital Sky Survey (SDSS). In order to gauge the clustering of the galaxies obtained from SDSS, the galaxies obtained from SDSS were compared to a distribution of galaxies whose members were randomly assigned a redshift, declination and right ascension within the ranges of the data obtained from SDSS. Next, the distance between all possible pairs of the galaxies from the SDSS galaxy distribution was calculated as well as the distance between all possible pairs of galaxies from both the SDSS data *and* the random-galaxies data. Using the distances found between pairs of galaxies denoted r, the number of galaxies separated by distances $r + dr$ was found for a range of r. Using the definition of the two-point correlation function described above, the correlation was then calculated.

Power Spectrum

The power spectrum of the two-point correlation function plays an important role in current cosmology. The relative peaks of the power spectrum provide constraints for any cosmological model of the universe. Mathematically, the power spectrum is the Fourier transform of the two-point correlation function:

$$P(\vec{k}) = \int \xi(s) e^{i\vec{k}\cdot\vec{s}} d^3\vec{s} = \frac{nV}{(2\pi)^{3/2}} \frac{|n_k|^2}{n^2}$$

where V is the volume of the galaxies, and n is their number density. By taking the inverse Fourier transform:

$$\xi(s) = \sum \frac{|n_k|^2}{n^2} e^{-i\vec{k}\cdot\vec{s}}$$

the power spectrum then has a clear physical interpretation. It is the amplitude of these "waves" of galaxy clusterings which the two-point correlation function characterizes.

Utilizing the above mathematical relations, the code for the power spectrum was straight forward. It is simply the Fourier transform of the two point correlation function.

Results

Overview

The purpose of the investigation of the large-scale structure of the universe was three-fold: (1) the investigation aimed to plot the large-scale structure of the universe with data obtained from SDSS; (2) a quantitative measure of the galaxy clusterings was sought; and (3) the investigation strived to calculate a power spectrum consistent with spectra already established in the literature.

Plotting the Large-Scale Structure of the Universe

Galaxies from SDSS were obtained and then graphed. Figure 1, below, contains a sample of galaxies with redshift $0 < z < 0.5$, right-ascension $0\,\mathrm{deg} < \mathrm{ra} < 360\,\mathrm{deg}$, and declination $-10\,\mathrm{deg} < \mathrm{dec} < 10\,\mathrm{deg}$.

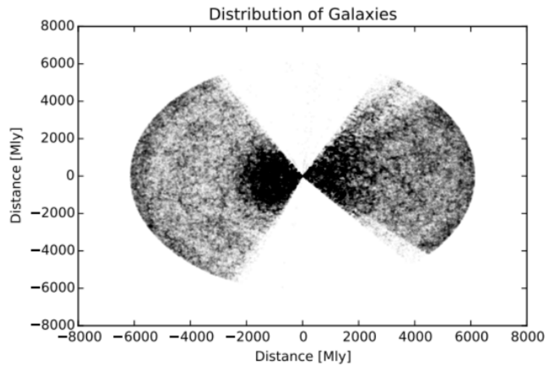


Figure 1: A sample of galaxies obtained from SDSS. The origin is chosen to be at Earth's center.

There are two observations to be noted from Figure 1. It is clear that the density of galaxies decreases the further one goes from Earth. If it is assumed that light in the universe has a relatively stable mean free path, then the probability that the light will collide and deflect from a galaxy is proportional to the galaxy's distance from Earth.

Another observation is the non-uniform distribution of the galaxies. If the plot above would be enlarged, the deviation from uniformity is very clear. Focusing on Figure 2, the structure of the galaxy distributions is revealed.

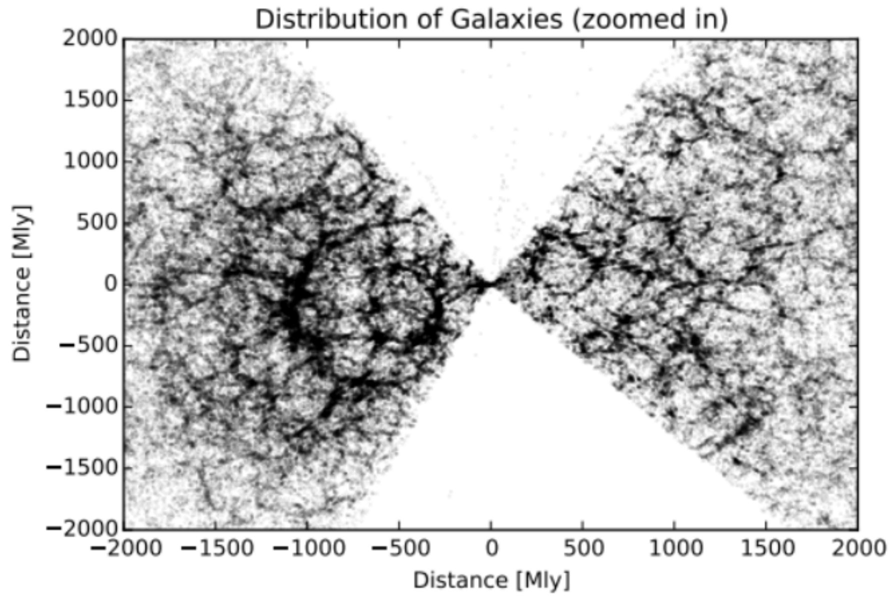


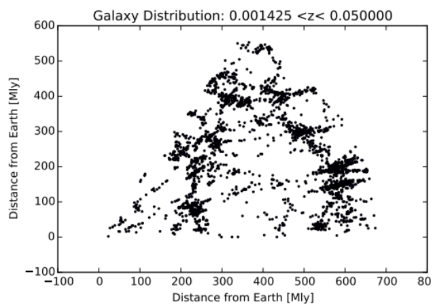
Figure 2: A zoomed look at Figure 1, highlighting the clustering / filament structure in the chosen area.

Particularly, viewing the left-side of the wedge in Figure 2, a semi-circular dark clustering is seen. This massive wall of galaxies is termed the "Sloan Great Wall". There also exist vast empty spaces with no galaxies at all.

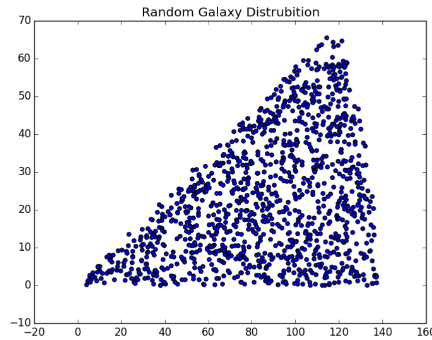
Thus, by plotting the coordinates of galaxies that were obtained from SDSS, the large-scale structure of the universe was revealed.

Galaxy Clustering

The next objective of the investigation of the large-scale structure of the universe was to quantitatively measure the clustering of galaxies appearing in Figures



(a) SDSS Galaxy distribution.



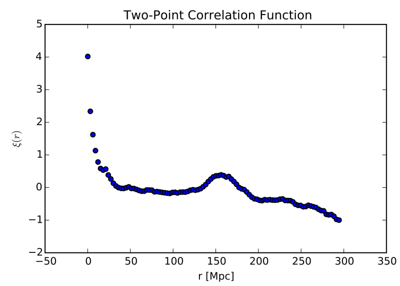
(b) Random distribution of galaxies.

Figure 3: Galaxy distributions obtained for their use in the two-point correlation function.$(0. < z < .01, -3.\,\mathrm{deg} < \mathrm{dec} < 3.\,\mathrm{deg}, 0.\,\mathrm{deg} < \mathrm{ra} < 60.\,\mathrm{deg})$

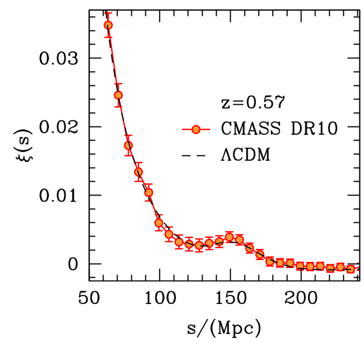
1 and 2. To measure this clustering, the two-point correlation function was introduced. In Figure 3, the galaxy distributions used in the calculation of the two-point correlation function are displayed.

In Figure 3(a) is the distribution of galaxies obtained from SDSS. In Figure 3(b) is the uniform distribution of galaxies obtained from simulation. Using these distributions, the two-point correlation function was calculated.

The correlation is seen to decrease exponentially as a function of distance. This means that it is relatively more probable for galaxy pairs to be closer than farther away. Therefore, the correlation coefficient is an accurate indicator of the clustering of galaxies consistent with the empirical data. Especially prominent is the bump in the correlation function at $r = 150\mathrm{h}^{-1}\mathrm{Mpc}$. The bump is invariant across all other published correlation functions as well, providing a test-case for our results. The bump has an important physical interpretation



(a) Calculated two-point correlation function



(b) Two-point correlation function of the BOSS DR10 CMASS sample.

Figure 4: The calculated two-point correlation function along with a published two-point correlation function.

for theories of cosmology. It is supposed that at the beginning of the early universe, quantum fluctuations created mass density fluctuations that propagated as sound-waves throughout the universe. Regions of galaxy clustering or lack thereof are indicative of these inhomogeneities in our early universe.

Power Spectrum

Given the bump in the correlation function, it appears that the correlation function has an oscillatory signal. To be sure, the Fourier transform of the correlation function was taken to reveal the amplitudes of this time signature.

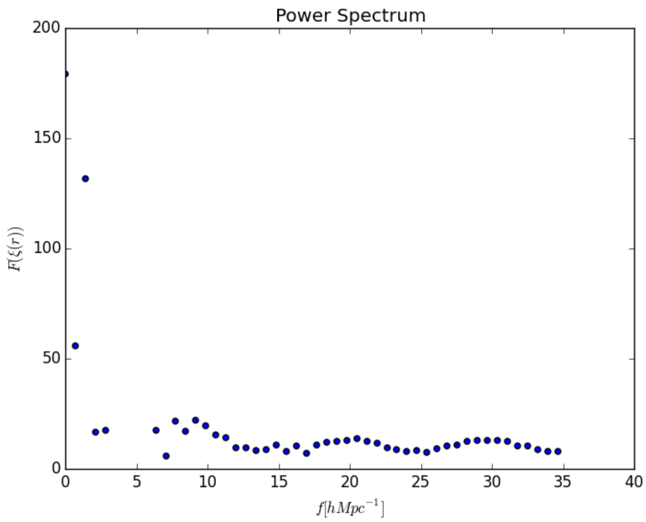


Figure 5: Power spectrum for distribution of galaxies obtained from the SDSS database.$(0. < z < .01, -3.\deg < \text{dec} < 3.\deg, 0.\deg < \text{ra} < 60.\deg)$

The power spectrum in Figure 5 is composed of a series of damped peaks. These peaks reveal the different modes of vibration in our early universe. However, although our results clearly reveal these small peaks, the literature expands the correlation function as a series of Legendre polynomials instead of as a series of complex waves. Qualitatively, though, our results are in agreement as Figure 6 reveals.

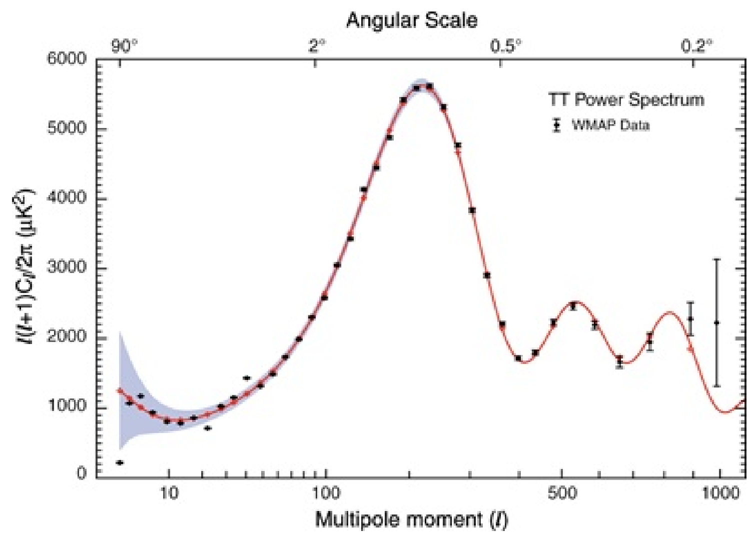


Figure 6: WMAP angular power spectrum: The first peak gives information about the total matter content of the universe; the second peak gives information about the total baryonic matter content of the universe [image credit:http://www.astro.rug.nl/ hidding/ao/ao.html]

Discussion

The purpose of this project was to investigate the large-scale structure of the universe. This was done by observing phenomenologically the large-scale structure; by mathematically characterizing the clustering between galaxies; and by finding the modes of vibration of these galactic clusterings within the universe.

To be able to obtain the location of actual galaxies in our universe and plot it on a coordinate space so as to reveal the large-scale structure of the universe was very satisfying. It was also very satisfying to create a suitable correlation function that clearly revealed the acoustic peaks of our early universe. Unfortunately, however, we did not project the correlation function onto the Legendre polynomials and so we were not able compare our results for the calculated power spectrum to the power spectrum found in the literature. Qualitatively, though, our results agreed: we were able to see the different modes of the acoustic peaks, albeit slightly.

Also, before beginning to plot the data, we learned that the median redshift of galaxy data collected by SDSS is 0.1, which corresponds to approximately 1,300 Mly. Due to the median, we expected to see a higher concentration around the 1,300 Mly point when the galaxies were plotted, however we were not sure how noticeable this would be until we plotted galaxy data of a redshift range from 0.0 to 0.5. This proved to be a very clear qualitative way of seeing the median redshift, and helped to showcase the results on the plot including all of the data. After we zoomed into the closer range, the expected filament structures were readily apparent, including the Sloan Great Wall as previously mentioned. This was a comforting check that the method we used to convert data was working.

The results of the two-point correlation function followed the general trend we expected before starting to create the function. We expected the function to decrease exponentially with a peak around 150 h^{-1}Mpc, and found that our two-point correlation function followed the decay and also had a peak close to 150 h^{-1}Mpc. Similarly, the power spectrum results followed the expected trend but with damped results. This makes sense as the power spectrum is the Fourier transform of the two-point correlation function.

Future Directions

If we had more time (and preferably more computing power), we would delve further into exploring the large scale structure. Our next step would be to find a way to generate the initial conditions of N galaxies. Ideally, we would write a basic N-body simulation code which could provide a loose model of the evolution of the galaxy positions over time and follow this up by running an established N-body simulation to compare our processes. Realistically this would be a difficult project for a future class to pursue. Alternatively, the results of the power spectrum could be used to generate various cosmological models. This could be an interesting way to compare the data with current models of the universe.

References

Information:

https://ned.ipac.caltech.edu/level5/Guzzo/Guzzo5.html

http://m.esa.int/Our_Activities/Space_Science/Planck/History_of_cosmic_structure_formation

https://www.sdss3.org/dr9/imaging/imaging_basics.php

Cosmological N-Body Simulations, J.S. Bagla and T. Padmanabhan,

http://arxiv.org/pdf/astro-ph/0411730v1.pdf

https://en.wikipedia.org/wiki/Sloan_Great_Wall

http://cosmic-horizons.blogspot.com/2012_02_01_archive.html

Equations:

The correlation function: Galaxies

https://ned.ipac.caltech.edu/level5/March04/Jones/Jones5_2.html

Accurate estimators of power spectra in N-body simulations, Stephane Colombi, Andrew Jaffe, Dmitri Novikov, Christophe Pichon,

http://arxiv.org/pdf/0811.0313v1.pdf

Examples/Diagrams:

http://naturalunits.blogspot.com/2012/12/dark-matter-iii-cosmology-evidences.html for examples of SDSS data, two-point correlation functions, and power spectra

http://inspirehep.net/record/1204394/plots for alternate looks at two-point correlation functions

http://wwwmpa.mpa-garching.mpg.de/galform/virgo/millennium/index.shtml looking at The Millenium Simulation

Code:

http://docs.scipy.org/doc/scipy/reference/tutorial/fftpack.html

http://physics.scsu.edu/~dms/cosmology/LSS_Lab.pdf

http://skyserver.sdss.org/casjobs/

Code

A current version of the code used for this project is available at:

`https://github.com/pitt1321/IndrasNet/blob/master/nbody_simulation/cosmic_web.ipynb`