MNRAS 000, 1Doc-Start–?? (2022) Preprint 12 September 2022 Compiled using MNRAS LATEX style file v3.0

Evolution of the large scale structure and stellar mass function

Anatoly Klypin[1,2]⋆,

Accepted XXX. Received YYY; in original form ZZZ

Key words: methods: numerical, galaxies: clusters: general, dark matter, large-scale structure of Universe, cosmology: theory

1 GOALS

Main goals of this project is to study the evolution of the large-scale structure distribution of galaxies and to analyze the stellar mass function. We start with the distribution of galaxies in space.

There are different statistical tools to analyze it including correlation function, power spectrum, void finders. However, those are complicated tools and in any case we need to start with learning what we are dealing with before we apply quantitative statistical measures.

Having this goal in mind, I made three plots showing first a large-scale view of the distribution of galaxies, then a zoom-in on a smaller section, and finally an even smaller section focusing on a massive cluster of galaxies. The pattern of galaxy distribution is quite complex. It has long connecting filaments (see Fig.1) with large objects – clusters of galaxies – found at intersections of those filaments (see Fig.2). There are large underdense regions of different sizes called cosmic voids. Typical size of those is $\sim 10-20h^{-1}$Mpc. The clusters of galaxies look quite spherical in their central $\sim 1h^{-1}$Mpc regions, but the pattern is more complicated once we look outside (see Fig.3).

Your goal is to study those features – qualitatively – at different redshifts. Find the most massive cluster in your sample at redshift $z = 0$. Make a large slice like one in Fig.1 with the slice going through the cluster. Now make the same slice at four redshifts: 0, 1, 2, 5. This makes nice 2x2 plot. Describe what do you see. Make two zoom-in plots centered on the cluster studying cluster environment and then cluster itself. Again, describe what you find.

Stellar mass function is another goal of the first project. Each object ("galaxy") in your sample has a stellar mass. The question is how the number of galaxies with different stellar mass evolves with time. We define the stellar mass function (SMF) as the number of galaxies with given stellar mass M per unit volume in a mass interval dM:

$$f(M,z) = \frac{dN}{V d\log M}, \quad (1)$$

where V is the volume of the sample.

Find SMF for five redshifts of your sample. Make a plot in log-log scales. Describe the evolution of SMF. On small masses $M < 3\times 10^{10}\mathrm{M}_\odot$ the stellar mass function is nearly a power law $f(M) \propto M^\alpha$. Estimate the slope. How does it evolve with the redshift?

What is the number-density and the average distance between galaxies *more massive* than $M = 10^9, 5\times10^{10}, 5\times10^{11}\mathrm{M}_\odot$ at different reshifts?

⋆ aklypin8@gmail.com

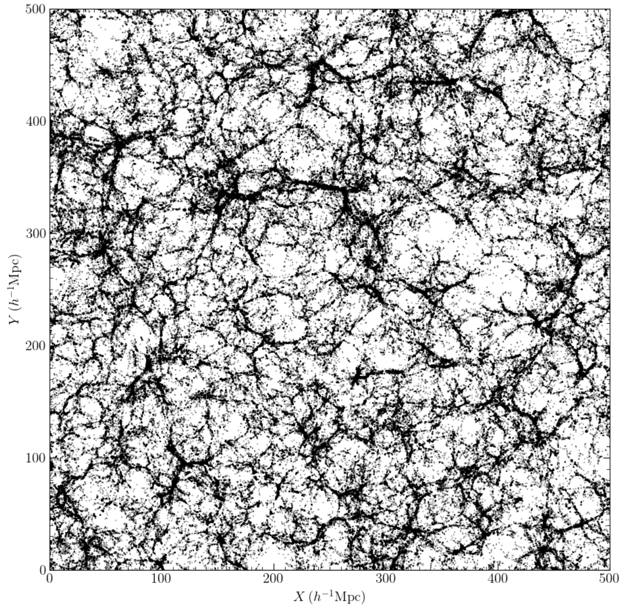

Figure 1. Distribution of galaxies in a $500\times500\times10h^{-1}$Mpc slice at redshift $z = 0$.

2 DATA

Data for all projects are on cosmology6 server. There are 12 directories named SampleXX (XX =01, 02,...12) that have the data. Each student must use only one directory. Names of files code sample number and redshift. For example file Sample.01.Redshift.0.000.h5 is for sample 1 and redshift $z = 0$. Each directory has data for 5 redshifts ranging from 0 to 5. At redshit $z = 0$ there are about 44 million galaxies and about 1 million at redshift 5.

There are 10 quantities that are given for each galaxy:

x, y, z – coordinates in comoving h^{-1}Mpc units
vx, vy, vz – peculiar velocites in km/sec units
SM – stellar mass in $\mathrm{M}_\odot$ units
SFR – star formation rate in $\mathrm{M}_\odot/yr$ units
$M_{\rm vir}$ – virial mass in $h^{-1}\mathrm{M}_\odot$ units
ID – index that tells whether object is central or satellite

Index ID is used to identify which galaxy is a central object and

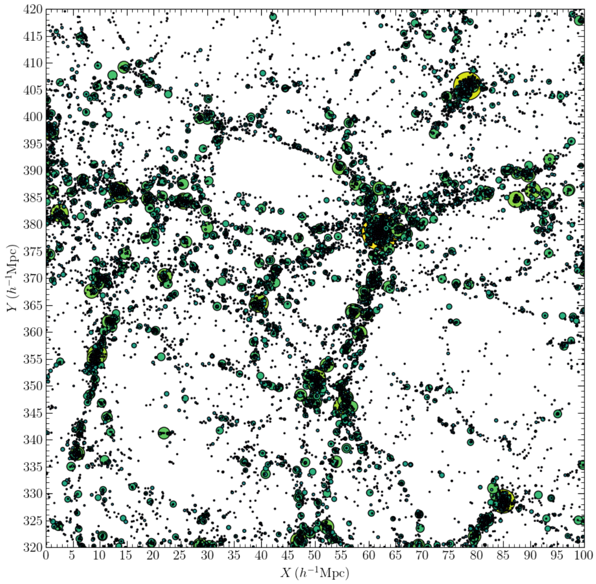

Figure 2. Distribution of galaxies in a $100 \times 100 \times 10 h^{-1}$Mpc slice at redshift $z = 0$. Radius of each circle is proportional to stellar mass of galaxy.

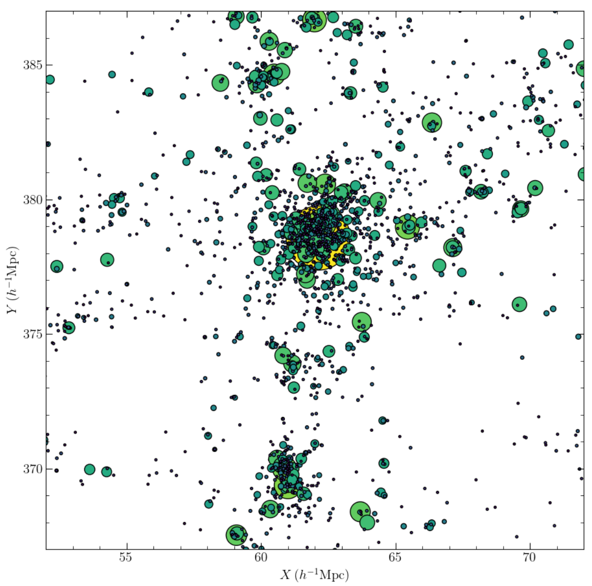

Figure 3. Distribution of galaxies in a $20 \times 20 \times 10 h^{-1}$Mpc slice at redshift $z = 0$.

which is a satellite. "Central" galaxy is the largest galaxy in a halo. For that galaxy one can define the virial mass. An example is our Milky Way (MW) and satellites around it. Virial mass of MW is about $M_{\rm vir} \approx 10^{12} \mathrm{M}_\odot$ and virial radius $R_{\rm vir} \approx 300$kpc. The stellar component is mostly confined to the much smaller inner $\sim$ 20kpc region where its mass is $M_{\rm star} \approx 6 \times 10^{10} \mathrm{M}_\odot$. Satellites by definition are galaxies within virial radius of the central galaxy. Satellites are typically heavily affected by the central galaxy and by dark matter halo. One normally does not use virial masses of satellites because satellites are stripped by tidal forces of the central galaxy, which drastically reduce masses bound to satellites.

Index ID of central galaxies is set to be ID=-1, and it is a positive number for satellites.

3 READING DATA

Data a written in hdf5 data format. Below are snippets of python how to read the data. One needs to import a library to read hdf5 files. This is done by routines in h5py:

```
import numpy as np
import math
import h5py
```

Open file Sample.01.Redshift.0.000.h5 and read some information:

```
h = h5py.File('Sample.01.Redshift.0.000.h5', 'r')
print("Keys: Box = h.attrs["Box"]
Ngalaxies = h.attrs["Ngalaxies"]
Redshift = h.attrs["Redshift"]
```

Here parameter Box is the size of the sample: $\mathrm{Volume}_{\rm sample} = Box^3$. For samples provided for the project $Box = 750 h^{-1}$Mpc though that can be changed. So, read it from the files. Parameter Ngalaxies is the number of objects in the sample. "Redshift" is the redshift. Now read all the rest of the data:

```
x = np.real(h.get('x'))
y = np.real(h.get('y'))
z = np.real(h.get('z'))
M = np.real(h.get('Mvir'))
SM = np.real(h.get('StellarMass'))
SFR = np.real(h.get('StarFormationRate'))

ID = np.int_(h.get('CentralObject'))
```