Application of Voronoi Tessellation on Finding Large-Scale Structures

Yongzhong Xu (T-8); ***xuyz@lanl.gov***

For astronomers to identify cosmological large-scale structures (LSSs), such as cosmological filaments, voids, galaxy clusters, and superclusters is not a trivial task. Although we have a large number of galaxy samples from the Two-degree Field Galaxy Survey (2dFGRS) and the Sloan Digital Sky Survey (SDSS), it is tough to find LSSs. Miller et al. [1] successfully identified galaxy clusters from the public SDSS Data Release One and Two (or DR1 and DR2) datasets, using their C4 algorithm. Hoyle et al. [2] identified galaxy voids from the SDSS DR1 dataset. However, no one has developed an algorithm to identify filaments or superclusters of galaxies from the available galaxy catalogs. In this writeup, we introduce an algorithm related to Voronoi Tessellation (VT), which plays a role, in principle, to find all kinds of cosmological LSSs.

As a test, we applied our algorithm on the Second Center for Astrophysics Galaxy Catalog (CfA2-GC). We successfully identified the known superclusters and rich clusters of galaxies in the nearby universe. Voronoi Tessellation and its geometric dual Delaunay Triangulation are robotic techniques which have extensive applications on various fields such as distribution of resources, cellular biology, statistics, galaxy distributions, and other fields related to the problem of closest neighbors. However, it is not an easy job to compute out Voronoi Diagrams (VDs) from a large number of random points. The quickhull (QHULL) algorithm [3] is very quick to derive the VT of a large number of random points in multidimensional space.

The CfA2 galaxy survey began in late 1980s. Its goal was to observe all the galaxies with b-band apparent magnitude brighter than 14.5 in the northern celestial sphere. The CfA2-GC was publicly released in 1998, which contains more than 18,000 galaxies with redshift measurements. Since this is a very shallow redshift survey and we know the large-scale structures in this region very well, we will use this catalog as a test for our algorithm of finding LSSs. First, we use the QHULL algorithm to tessellate the CfA2 Galaxy Catalog in 3-D space. We obtain a well-defined data structure that contains detailed information of every VD, for instance, the neighbors, the vertices of each face, and the norm vector of each face. Using this information, we calculate the physical volume of each VD, and we assume that these volumes of VDs are the "volumes" occupied by galaxies.

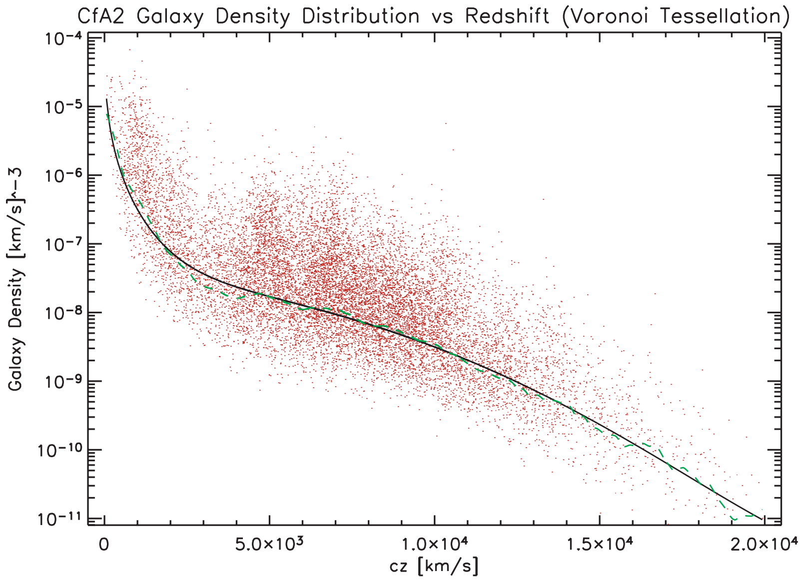

Figure 1*—*
Galaxy number density distribution as a function of redshift in the CfA2 Galaxy Catalog. The green dashed curve shows the mean galaxy number density, and the solid curve shows the best-fitted mean density.

Therefore, the number density of a galaxy is just simply the inverse of volume of the VD of this galaxy. Those scattered points shown in Fig. 1 describe the galaxy number density as a function of redshift. Given a redshift bin, it is trivial to calculate the mean number density of galaxies within this redshift bin. The green dashed line is the mean number density that we calculated from the real data, and the black solid line is a best-fitted curve of the mean number density. We calculate the overdensity or the contrast of number density of every galaxy. Of course, we need to remove those galaxies near the edge of the survey, those near the equator of the Milky Way, and those whose Voronoi Vertices are out of the survey range, since their volumes are overestimated.

Second, just getting the VT and the overdensities of galaxies does not mean that the job is done. We need to search for structures with similar properties (such as overdensity and color). The basic idea of searching for structures is to find the adjacent VDs having the given overdensity limit. We developed a quick algorithm searching for structures of galaxies with overdensity inside a certain range, based on the special data structures from QHULL. Our algorithm has a theoretical speed of O(M*N*LOG2(N)). It is worthy to point out that the number N in the previous expression is the number of galaxies that are within the overdensity limit, and the number M is the total number of structures. This number N is, in general, much less than the total number of galaxies. Astronomers believe that overdensity inside clusters is significantly higher than that inside filaments, and that overdensity inside filaments is significantly higher than that inside voids of galaxies. However, we do not know the overdensity threshold of galaxy clusters, nor that of galaxy superclusters either. After carefully testing with various limits, we find that overdensity larger than 2.0 is a reasonable limit for searching for superclusters in CfA2-GC, and that overdensity larger than 20.0 is a reliable limit for clusters, and that overdensity less than -0.6 is a possible limit for voids of galaxies. Figure 2 shows the nine superclusters that we found in CfA2-GC. The figure is plotted with Aitoff projection in Galactic Coordinates. The nine superclusters are plotted with nine different colors respectively, including the four famous nearby superclusters: Perseus-Pisces Supercluster, Hercules Supercluster, Virgo Supercluster and Coma Supercluster. We still need some further work to obtain the void catalog.

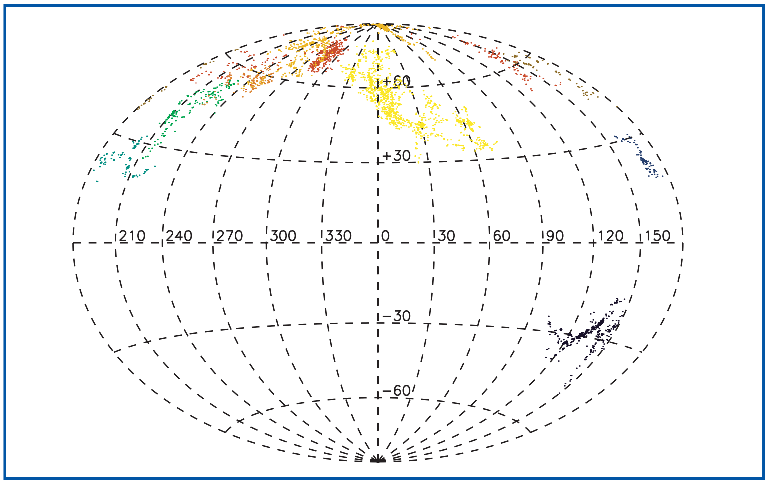

***Figure 2—* Nine galaxy superclusters are plotted in Aitoff projection in Galactic coordinates. Different superclusters are shown in different colors respectively.**

After a lot of tests with CfA2 Galaxy Catalog, we claim that our LSSs finder is robotic to find any kind of large-scale structures, especially to find cosmological filaments or superclusters of galaxies. Currently, there are several large-area galaxy surveys available in the astrophysics community. The Two Micron All Sky Survey (2MASS) has successfully observed the whole nearby universe. Although it is just an imaging dataset, we can still use it to find some interesting unknown structures, especially in the southern sky, since it observed the infrared universe. The Sloan Digital Sky Survey (SDSS) has just published its third Data Release (DR3), including both photometric and spectroscopic data. The Two Degree Field Galaxy Redshift Survey (2dFGRS) has finished its observation on a deeper universe than the other two surveys. By applying our algorithm on those datasets, we believe that we can discover a lot of interesting LSSs.

[1] Miller et al., SDSS internal paper.
[2] Hoyle et al., Astro-ph 0309728.
[3] Barber, Dobkin, and Huhdanpaa, *ACM Transactions on Mathematical Software* **22**, 4 (1996).