Improving the Voronoi Tessellation Cluster Finder Algorithm

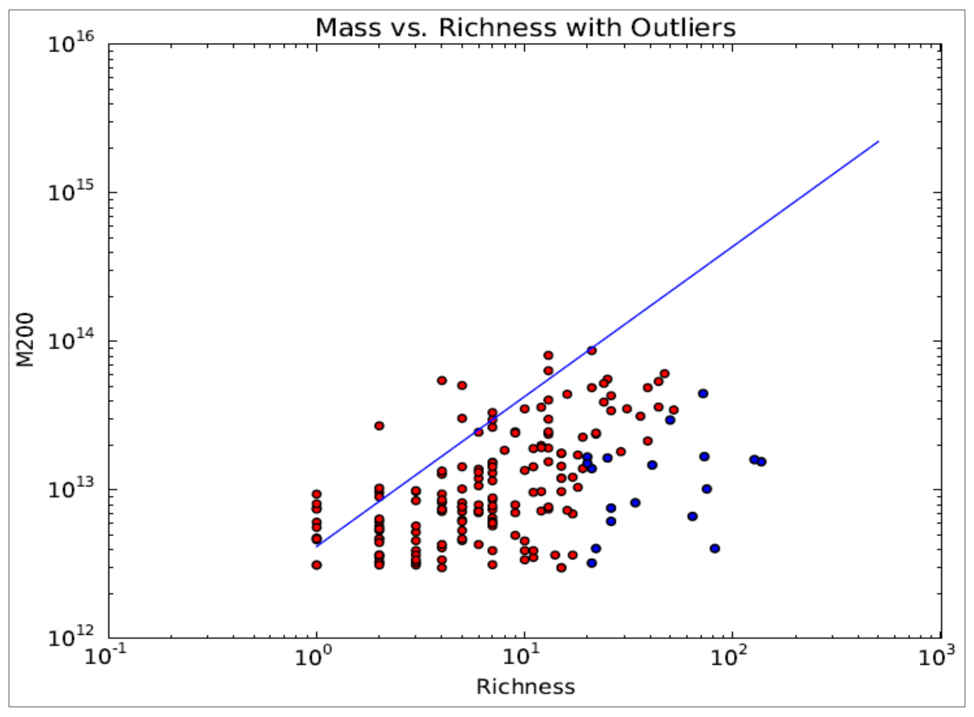
Tim Osborn, SULI Program, Fermilab

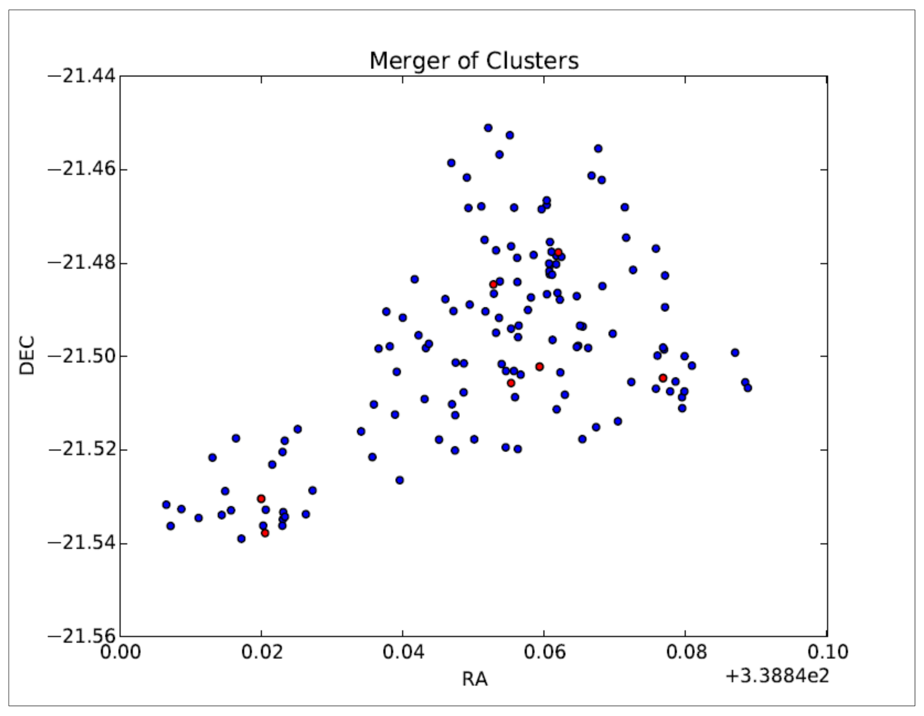
MOTIVATION

In order to study the mechanisms that govern our universe and its expansion on a cosmological scale, we must use the largest gravitationally bound systems in the universe: galaxy clusters. While these objects are straightforward in definition, a group of galaxies that are near each other and gravitationally bound, finding them can be a tedious and difficult process, with many challenges to overcome requiring creative solutions. The Voronoi tessellation (VT) cluster finder algorithm is one such solution that relies on the use of Voronoi tessellations in order to determine if a group of galaxies is in fact a cluster or not. Using data collected from the Dark Energy Survey, this algorithm can create a comprehensive catalog of clusters including their location in right ascension (RA) and declination (DEC), their redshift, their richness or number of galaxies contained within them, their mass, and other characteristics. While this algorithm does a great job finding clusters and cataloging them, improvements can always be made. The goal of this project is to find weaknesses within the VT cluster finder algorithm and implement solutions to minimize the impact of those weaknesses.

METHODS

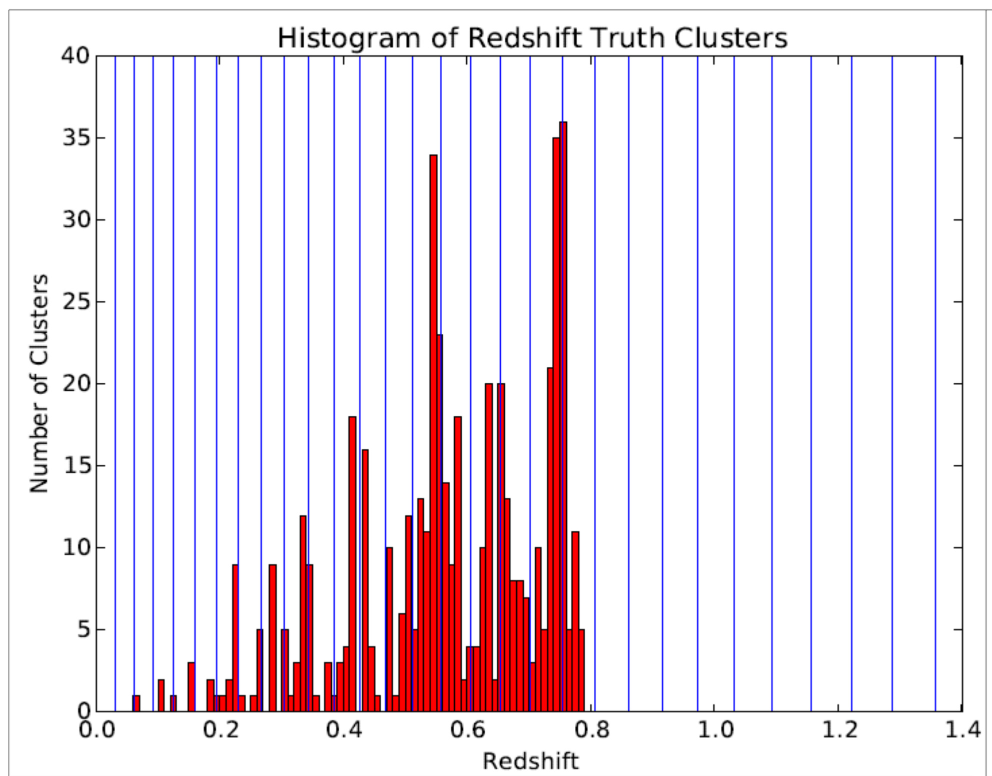
In order to find the weaknesses within the VT cluster finder algorithm, it is necessary to run the algorithm on data from simulations of the universe. This allows for the analysis of outputs from the algorithm against “true” results, or the known conditions of that simulated universe. This analysis takes the form of multiple plots that allow for visualization of discrepancies between the “true” conditions and the output of the algorithm. The plots used most often include: histograms of redshift, scatter plots of mass vs. richness, and scatter plots of RA vs. DEC. Two examples of these plots are shown below. From this visualization of the data, it becomes clear when a potential weakness is exposed. As that problem is uncovered, tests can be run by making subtle changes to the algorithm and repeatedly analyzing the results to see if improvement has been made to the problem. This process is repeated until a method is determined that can minimize the effects that this weakness or problem has on the efficiency of the algorithm as a whole. The graphs below demonstrate the problem of true clusters being merged into one larger cluster.

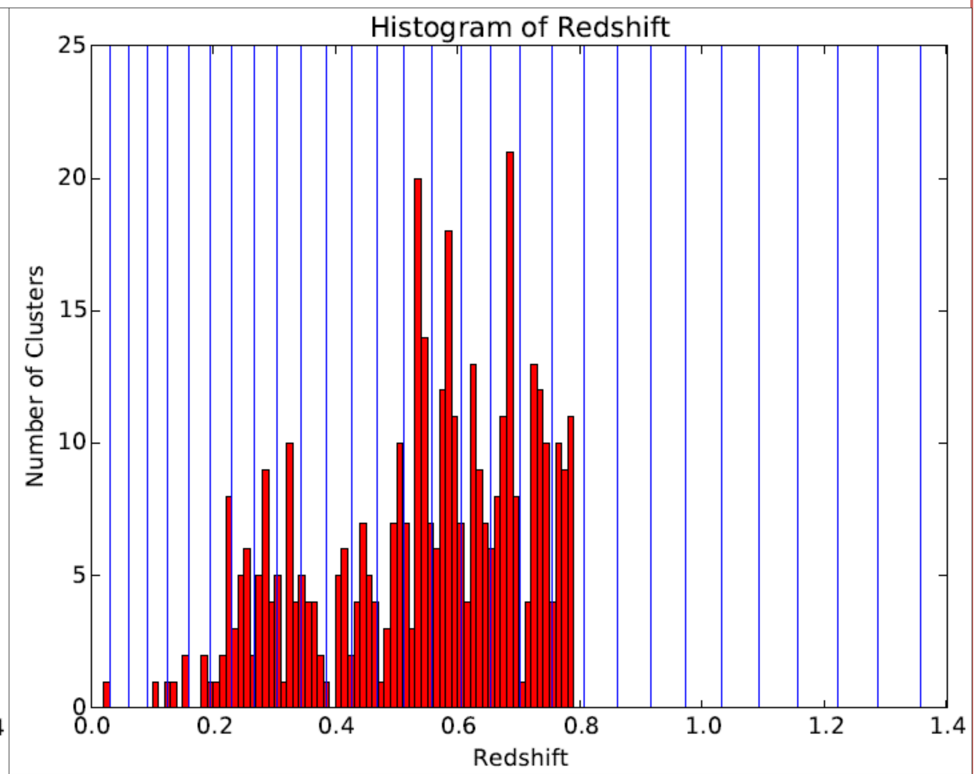


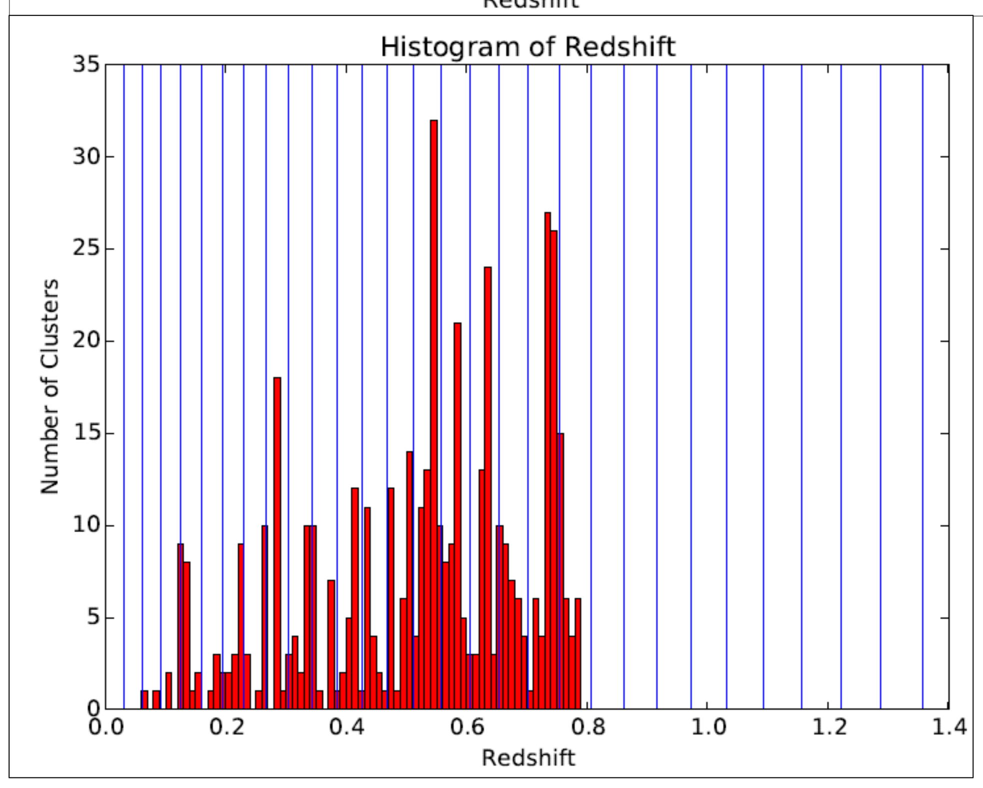


RESULTS

Through analysis conducted two key weaknesses were highlighted: the merging of multiple small true clusters into one large cluster and the loss of clusters that fell on or near the redshift boundaries set by the algorithm. Visuals of these problems are shown below:







Top Left Plot: True distribution
Top Right Plot: Original output distribution (Note lack of clusters along boundaries.)
Bottom Left: Output that used a single defined redshift (from the truth) for galaxies that should appear in the same cluster. Clusters are now found on boundaries, demonstrating that the problem was due to a splitting of galaxies that should be in the same cluster by the redshift boundaries.

SOLUTIONS

Through using the diagnostic plots as a guide, it was possible to determine the cause of one of the major problems found with the algorithm, the loss of clusters along the set redshift boundaries, and fix that problem. It was determined that the cause of that loss was that because the galaxies in a single cluster will all have slightly different redshifts, it was possible for there to be some galaxies on either side of the redshift boundary. This caused the cluster to be split into smaller sections by the algorithm and, unfortunately, smaller clusters are often missed. The solution to this problem was to keep the boxes of redshift approximately the same size but use overlapping redshift boundaries instead of the rigid sections that the algorithm had used initially. This created a greater chance of finding the entire cluster within one of the boxes of redshift. With a greater chance of finding these previously discarded clusters, the algorithm now creates more complete catalogs of galaxy clusters, allowing for better analysis of the data collected from surveys.

August 2015

Fermilab

U.S. DEPARTMENT OF ENERGY | Office of Science