Taxonomic Homogeneity of Asteroid Clusters

Zachariah Germain, ZachGermain, zgermain@udel.edu

Abstract: I clustered asteroids using a k-means algorithm based on their orbital parameters in order to determine if spectral type, specifically SMASII taxonomy is homogenous within asteroid clusters. If that was the case, then using the essential orbital features, one could predict the spectrum of an asteroid and with that, its likely composition and structure, leading to a greater understanding of the asteroid belt. My analysis does not support the conclusion that asteroid clusters are taxonomically homogenous, thus the orbital parameters cannot be used for spectral type prediction. But, to some degree the reverse conclusion is supported, that some spectral groups lie within a select range of orbital parameters. This could open up other avenues of questioning.

Introduction: It would be a great step in asteroid science if there was an easier way to determine asteroid spectra. Currently it is difficult with surveys like SMASSII (Small Main-Belt Asteroid Spectroscopic Survey) which was only able to determine spectras of about 1400 asteroids of a population of around 800000 individuals[1]. Current methods of asteroid clustering center around discovering asteroid families. These are taxonomically homogeneous, but mostly by definition. They are so because the current theory is that they come from the same collision and as such have similar compositions and structures. They are additionally clustered and determined based on non-orbital parameters so nothing can be determined based off of their orbits and many known interlopers (asteroids who have orbits within a family range but are so due to orbital migration) exist. Another issue is that orbital evolution is unpredictable above a million years, preventing the usage of family detection based on orbits alone for most asteroids. As of now, less than 10% of asteroids in the Main Belt have been identified as members of dozens of families[2]. Thus, it would be a huge advantage if asteroid spectra could be determined indirectly and without the reliance of the large number of orbital and non-orbital parameters required for family identification. With this analysis, I made one step towards determining if the spectral type of an asteroid could be identified using parameters of an asteroid that are easier to measure, such as the major features of their orbit, a (semi-major axis), e (eccentricity) and i (inclination of orbit). One way to determine this is to cluster the asteroids by these orbital parameters and determine their purities. If they have a high purity, then it is very likely that the spectral type of the asteroid can be determined by a known range of orbits.

Data: I acquired the data from the JPL Small-Body Database Engine[3]. I exclusively selected asteroids that were members of the Main Belt (2.0 au < a < 3.2 au and perihelion distance > 1.666 au). I also only used asteroids with defined SMASSII spectral types, semi-major axis, eccentricity and inclination. I used their internal asteroid identification system as my indices for the dataframe. I then reduced the data, getting rid of all asteroids of spectral types containing less than or equal to 20 asteroids as these would be small enough to be outliers. 20 asteroids is about 5% of the number of asteroids of the largest spectral type, S. I then had 1105 asteroids split between 14 spectral types. I scaled the three parameters to be a normal distribution and checked they were independent of each other. This is clear in Figure 1, a scatter matrix, It shows areas of high and low densities, but no clear covariance is exhibited so they can be useful in clustering. Another view of the data comes in the form of the 2D TSNE projection of the orbits, as in Figure 2.

Methodology: The central part of this analysis is the clustering. I chose to use the same number of clusters as there are spectral types. This gives me a way to determine if spectral groups are centralized, as well as determine if there exists a one-to-one correspondence between orbit range and spectral type. Using a larger number of clusters would allow for the ability to detect clumped distributions spread across the parameter space but runs the risk of overfitting. I then chose the cluster centers I wanted to use based on the centers of the spectral groups. This again allows for the detection of spectral centralization. The alternative, an ensemble method of k-means clusters with n-iterations was also tested, but it comes out with very similar results and fails to have the physical meaning of the spectral group centers. With this, I can identify the clusters, as in Figure 3. In order to discover if the clusters are homogenous, I should look at the purities of each cluster, a heatmap is provided in Figure 4. S-types are the

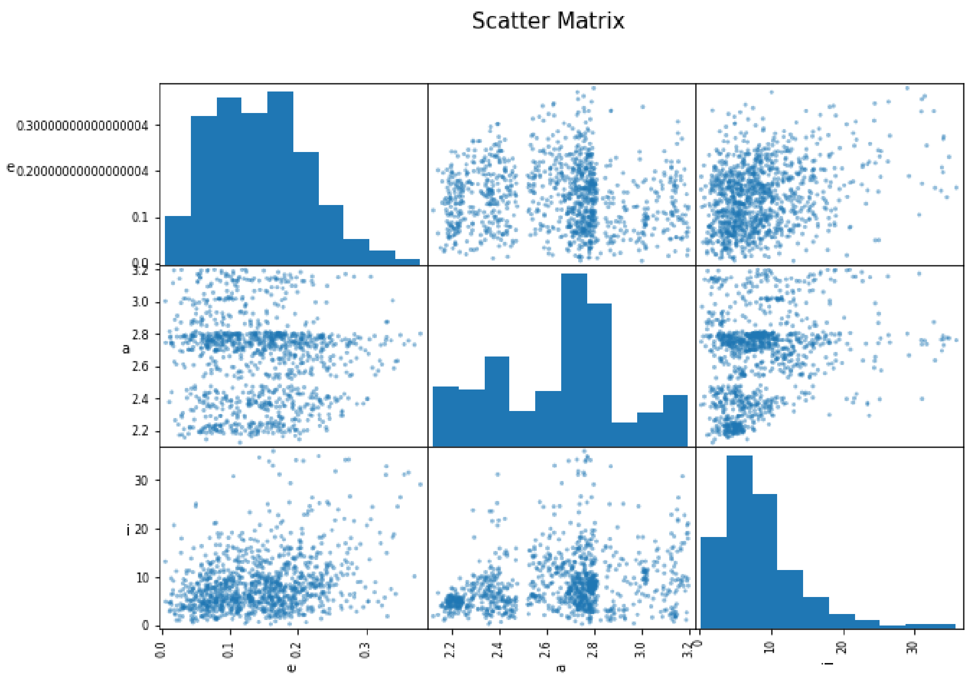

Figure 1. A scatter matrix of the orbital elements. There are regions of high and low densities in each graph in the figure, but it does not seem to result in covariance.

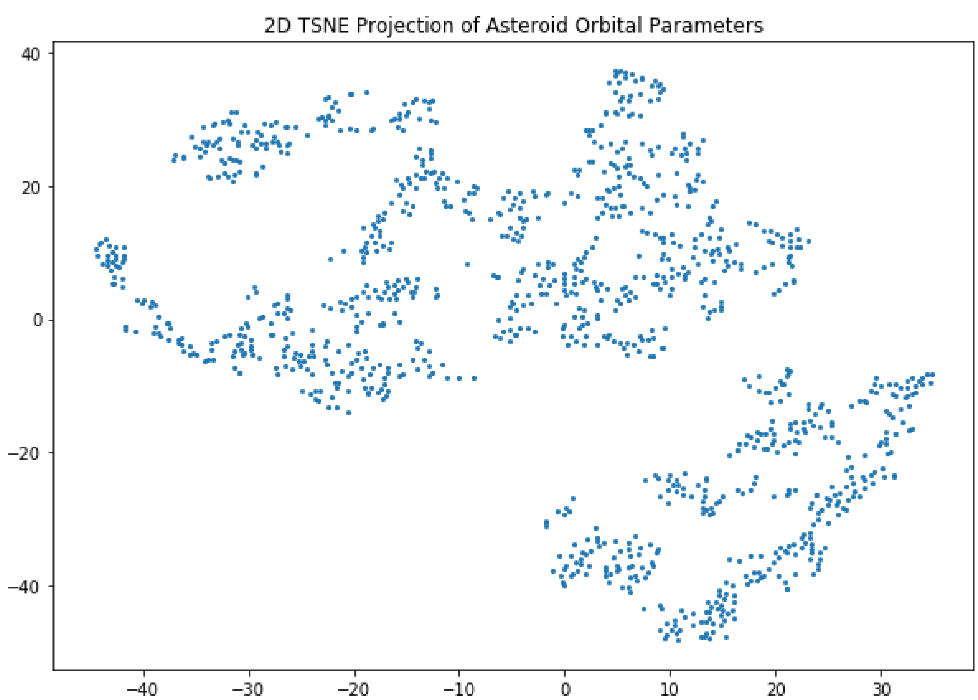

Figure 2. A 2D TSNE projection of the parameter space. This also shows large gaps in the parameter space, and regions of high density, promising for clustering efforts.

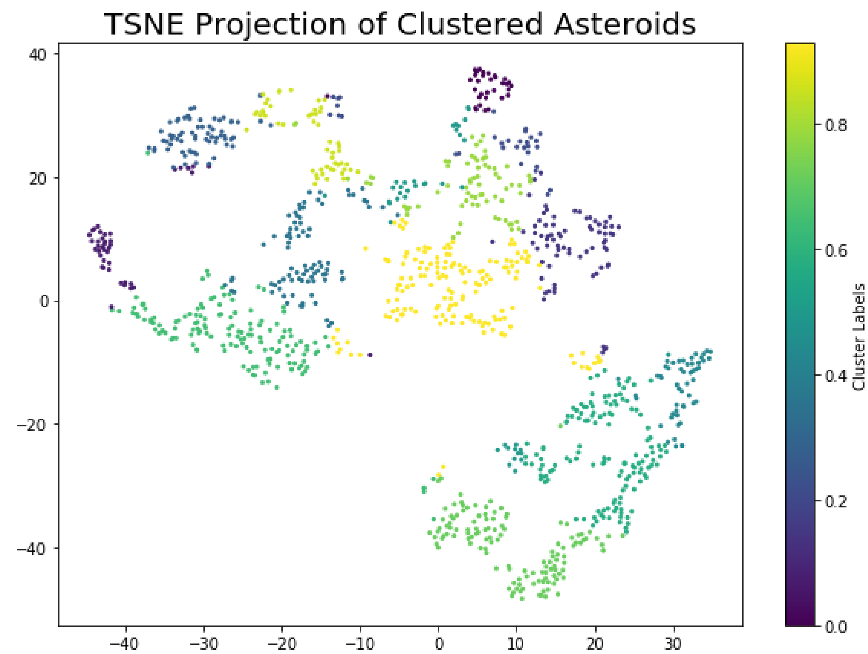

Figure 3. A TSNE projection of the clustering. There are 14 clusters observed, each largely confined to a section of the parameter space as one would expect.

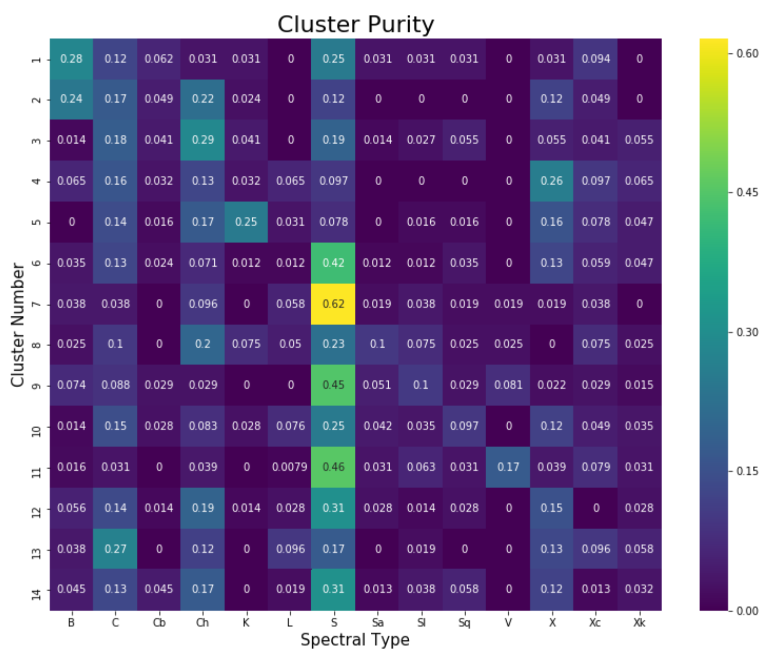

Figure 4. A heatmap of the percent compositions of clusters. The more yellow, and less purple spectrals of a cluster are the purities. S-type asteroids clearly dominate the clusters, consisting of 31% of the population and being the spectral type associated with the purity of 8 clusters.

most common purity type, with it being the type of 8 clusters, only 4 of which they are a larger percent composition in the cluster than in the population, 31%. S-types are also the type of all 4 clusters above 100 individuals. B-types are the types of 2 clusters, but their purity is under 0.30, consisting of 5% of the population. Both of those clusters are smaller, around 40 asteroids in size. The other 4 clusters have different purity-associated types and whose purities are below 0.30. This information can be found in Figure 5.

Conclusions: This analysis fails to support the idea that asteroids can be clustered based on orbital parameters in ways that would allow prediction of their spectral types. It seems that many groupings of asteroids based solely on their orbits leads to spectral type distributions that have ranges too large and overlapping to allow for spectral type prediction. Though many have peaks, there is too much interference for precise determination. The distributions of spectral types however, show results of greater use. Several types are concentrated in clumps within clusters; K and V types being the best examples (Figure 6). Many are also fairly evenly spaced across the Main Belt. The concentration of some types imply they have been created in only a few select ways. It also means that we could discover what orbital parameter ranges these types exist in.

Future Work: Potential ways to add to this analysis would be to remove outliers, try other clustering methods or use decision trees. You could also look at other regions of the solar system. To add to the parameter space, other non-orbital parameters could be added such as albedo or color while attempting to minimize required information and maximizing output information. Another avenue would be to look further into the clusters containing spectral clumps and attempt to isolate those groups to see if they are spread evenly within those clusters or even more concentrated than we can determine by this analysis.

Cluster Number	Cluster Size	Cluster Purity	Spectral Type
1	32	0.281	B
2	41	0.244	B
3	73	0.288	Ch
4	31	0.258	X
5	64	0.250	K
6	85	0.424	S
7	52	0.615	S
8	40	0.225	S
9	136	0.449	S
10	144	0.250	S
11	127	0.457	S
12	72	0.306	S
13	52	0.269	C
14	156	0.308	S

Figure 5. The table of information on clusters, their purities, spectral types and size.

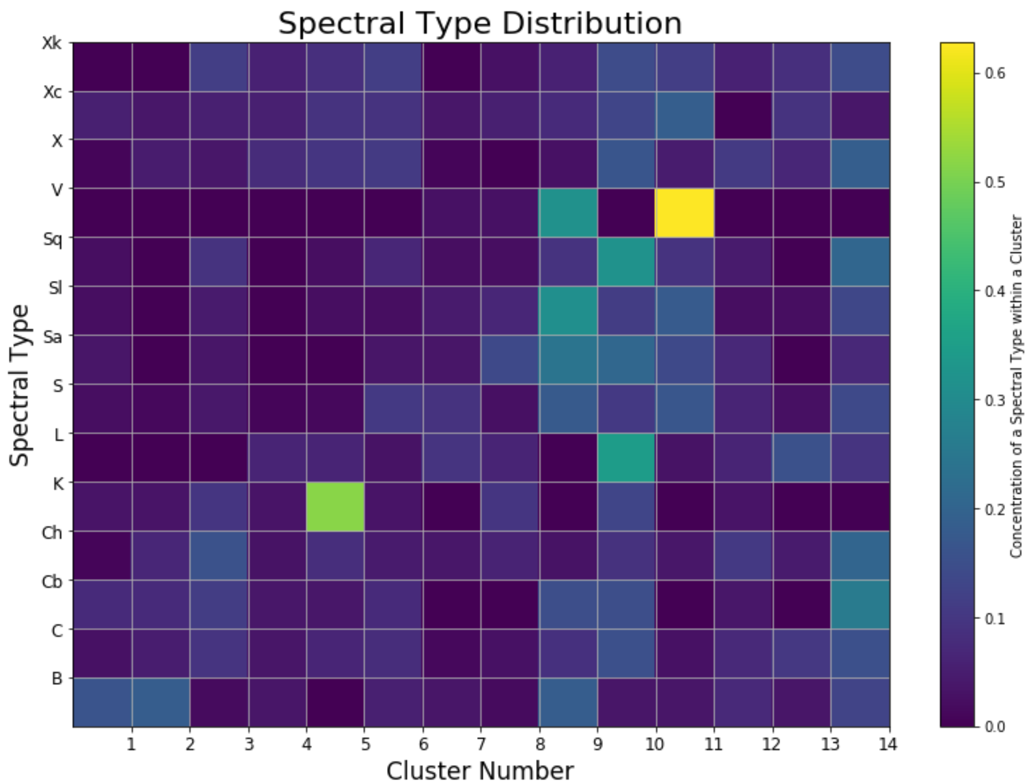

Figure 6. Heatmap of the spectral type distributions. As can be seen, some types like V, Sq, Sl, L and K have a significant proportion of their spectral group in a single cluster.

Links: Github Repository Code Folder Data Folder

Bibliography:

1. DeMeo, et al. “The Compositional Structure of the Asteroid Belt” Asteroids IV, University of Arizona Press, 2015, pp. 13-41
2. Nesvorny, et al. “Identification and Dynamical Properties of Asteroid Families” Asteroids IV, University of Arizona Press, 2015, pp. 297-321
3. “JPL Small-Body Database Search Engine.” *NASA*, NASA, https://ssd.jpl.nasa.gov/sbdb_query.cgi#x.