Capstone Project 2 Proposal - Dan Feldman

Discovering Stellar Clusters with Gaia

This project seeks to discover and better flesh out stellar clusters in the nearby neighborhood of the Milky Way Galaxy. Our galaxy, the Milky Way, is a barred-spiral galaxy 100,000 light years across, containing roughly 100 billion stars. Within our galaxy are a large number of stellar clusters and moving groups, which are numbers of stars that have about the same age and distance, as they formed from the same cloud of gas. Understanding these clusters can not only illuminate the population of stars within them, but also when studied in aggregate, the evolutionary history of the galaxy itself.

Astronomers have used a variety of techniques to discover and populate clusters and moving groups of stars; often, however, they've had to take data from a number of telescopes accumulated over a large period of time to slowly build catalogs for clusters. That is because cluster membership requires knowing stars' 3-D positions in the galaxy and their motion on the sky, a very difficult task for all but the nearest stars. Now, the ESA telescope Gaia, launched in 2013, has been cataloguing the data necessary to obtain these data for over a billion sources, stretching a third of the way into the galaxy or greater. As of their recent second data release (DR2), they have all 5 necessary features for over 1.3 billion stars.

I will download the DR2 catalog, which is open to the public, in order to try and discover new stellar clusters/moving groups, and to potentially add members to already known clusters. I believe this can be done using unsupervised learning clustering techniques, which is a natural direction to guide such an endeavor. Given the massive size of the data, I will use techniques for machine learning at scale, like running a databricks instance with Spark on Microsoft's Azure cloud services.

The hypothetical client for this work would be the astronomy researchers and/or institutions that could use this work for their research into stellar populations and galaxy evolution. Having a way to determine cluster membership would be very important for their research, leading to a number of academic papers and the possibility of additional funding for the institutions and research groups. Additionally, as more data from Gaia (and other future missions) is released, these techniques can be re-applied to continue this type of clustering.

My deliverables for the project would be the code, tables and plots describing various clusters discovered/populated for research groups/institutions to use, and a paper describing the findings.