

# Project Proposal: Identifying outliers in the latest quasar catalog

Reza Monadi, Hasin Us Sami, Arman Irani, Terrance Kuo

October 15, 2020

## Project Summary

Given a catalogue of 750,414 quasars, we will use four different outlier detection methods to identify those quasars that are very different from the majority of quasars' population. Finding these outlier quasars gives an insight into future research opportunities by proposing a modification for the current astronomical models for quasars.

## Scientific Background

Quasars are a class of galaxies with a super massive black hole at the center. The black hole devours surrounding materials and converts them to electromagnetic radiation in different wavelengths as a 1D image which is known as *quasar's spectrum*. This spectrum has some absorption and emission features that may be different throughout quasars due to various physical conditions.

## Scientific Justification

The main purpose of this project is to find those quasars which are less similar to the typical normal quasars. These outliers can be simply some objects that are mistakenly classified as a quasar in the main catalog or they can be the representative of a class of very interesting objects that current astronomical models cannot explain. In either cases identifying outliers will be beneficial. We are going to use a variety of outlier detection algorithms for identifying outliers. In this case we can compare the results of each algorithm and be confident when more algorithms are consistent about a particular quasar as an outlier. Because the overall shape of the 1D image of a quasar (spectrum) depends on its color, we separate quasars first based on their colors and then identify outliers in each class.

## Data

The latest quasar catalogue dr16q includes lots of information about individual quasars obtained by analyzing their spectra. We also have access to the spectra

of individual quasars in dr16q publicly available in Sloan Digital Sky Server webpage SDSS

## Evaluation

- There are previous outlier analyses performed on former quasars catalogs. We can compare our results to them.
- Outliers in each class should be visually inspected by looking their spectra to see why they have been labeled as outliers. We do this by stacking spectra of a group of outliers which are closer to each other and then compare it with the median spectrum of a numerous normal quasars in the catalog.
- The output of different algorithms we are implementing here might be different. This let us to compare the performance of each algorithm.

## Labour division

The data set has 750,414 quasars with 183 features. However, we will use a subset of relevant features. In the pre-processing we select most important features and then normalize the data. Afterwards, we will apply the following outlier analyses:

- **Hasin Us Sami:** Implementing Density-Based Spatial Clustering to detect outliers.
- **Arman Irani:** Isolation forest
- **Terrance Kuo:** Agglomerative Hierarchical Clustering
- **Reza Monadi:** Implementing Kohonen self-organized maps to reduce the dimentionality of data to visualize the identified outliers.

## Relation to Reza Monadi's thesis project

One of the projects in my thesis is about studying extremely red quasars (ERQs). The main purpose of ERQ project is to see if ERQs are outlier or just extreme cases of the main population. Our data set here, however, is different and the type of outlier we are going to find is not known before hand. Moreover, the physical parameters which build our feature space is different here. This project can help my thesis if we find some outliers that happen to be ERQs.