

Dustin Burnham
CSE 160
HW 7 Part 2
3/10/17

Circumgalactic Gas Flows that Drive Galaxy Formation and Evolution
By Dustin Burnham

Summary:

1. By analyzing the gas between the milky way and reference light source like a quasar, we can see heavy elements at a certain distance that are a part of a galaxy's CGM. How far does the circumgalactic medium (CGM) extend from its host galaxy?

-By using cosmology math, we can determine the distance between us and the two objects, and then with trigonometry find the actual distance between the two objects. My results that I found were that CGM can range between 46 kpc and around 900 kpc. I was able to map each system of strong hydrogen to a "nearby" galaxy, except for one. This outlier is caused because the uncertainty in the host galaxy is very large compared to its distance.

2. What can the composition of these CGM tell us about galaxy evolution and dynamics?

-We can look at the metals found in a Galaxy to tell us about what is currently happening in that galaxy. Is it currently forming stars (young or old) and what is causing the CGM to be thrown out to these distances? The more metallic a galaxy, the older the galaxy is, and by looking at the results I calculated, a number are old and some are younger. Further analysis of these properties will give more exact answers to these questions.

Motivation and Background:

I do research in the astronomy department. The reason I chose this topic is because this is what I find interesting, it's important to modern astronomy, and I thought it would be cool if I could tie my new knowledge of python to help answer questions from my datasets. In my research group, I analyze quasar spectrum, from the Hubble Space Telescope: Cosmic Origins Spectrograph, searching for systems of gas between the milky way and the quasar. The question of size of the CGM can tell us how gas flows in galactic structures, and thus help make clear how this affects galactic evolution. This is interesting because this is a semi-new topic in galactic astronomy, and will continue to become more important as more QSO spectrum are analyzed in the future.

Dataset:

I will use two data sets to calculate the lower limits of the CGM. One is the data that I have analyzed through research that is stored in a .json file, and the other is a data on website. I have analyzed quasar spectra with a Python GUI, saving the data in a .json file. This .json file

contains the info of the identified elements like the element name, distance, column density, comments, rating, and others. Currently, there is some amount of code provided to my research group to make the data more readable, by turning the data into a data set organized by distance. I would like to pivot this data to be organized and filter the data by column density, giving good locations to look for galaxies. The filtered data will then be turned into a dictionary within a dictionary, with the key being the system and the value being a dictionary containing all of the info at that location.

This leads me to my second data set, which is a list of objects from the SDSS DR7 website. This website contains information about galaxies in the general area of our data set. Here we can locate galaxies and look at their distances measured in redshift, the uncertainties in the distance, the galaxy ID, and the coordinates. Using a SQL query, I was able to cross reference two data sets. This was needed because the redshifts and redshift errors needed did not include galactic coordinates, so I had to reference both PhotoObjAll and Photoz datasets saving the result to a plain text file. The range of sky that I searched was a 0.5 degree square box around the QSO. Using these data, I cross referenced these redshifts with the redshifts of systems of hydrogen. At this point I made calculations to find the angular distance and actual distance, and make plots of the data from the coordinates of the matched pairs.

Query Search Used:

```
select
PhotoObjAll.objID, PhotoObjAll.ra, PhotoObjAll.dec, Photoz.z, Photoz.zErr

from
PhotoObjAll, Photoz

where
PhotoObjAll.objID = Photoz.objID AND
PhotoObjAll.ra BETWEEN 153.909418 and 154.19418 AND
PhotoObjAll.dec BETWEEN 47.01204 and 47.21204
```

Galaxy Data from SDSS query:

```
objID ra dec z zErr
245953888996294656 153.1173 47.62363 0.162356 1.33051E-4
245670976484802560 153.1175 47.569511 0.162939 1.83449E-4
245670974278598656 153.12111 46.884929 0.165449 1.07673E-4
245670974035329024 153.13481 46.695356 0.038186 9.47037E-5
```

Sample of analyzed Data

```
{
  "bad_pixels": [],
  "cmps": {
    "z-0.00001_CI": {
      "A": null,
```

```

"Comment": "MW Cl",
"DEC": 0.0,
"Ej": 0.0,
"Name": "z-0.00001_Cl",
"Nfit": 16.3,
"RA": 0.0,
"Reliability": "a",
"Zion": [
    6,
    1
]

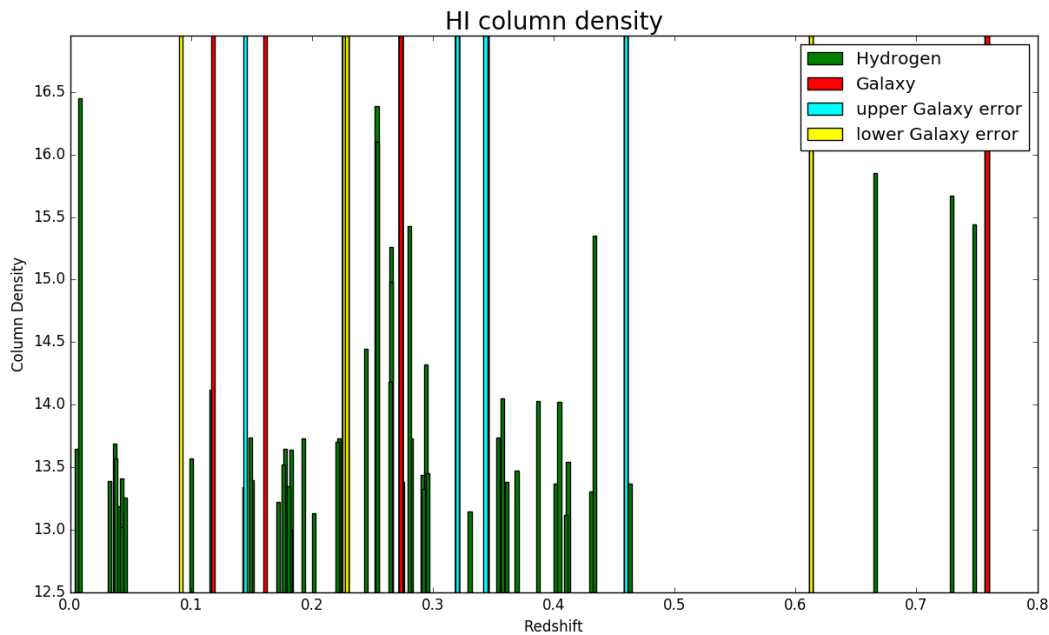
```

Methodology:

Start by reading in a .json file that contains the data. The data is obtained by analyzing a spectrum of a quasar using a python GUI that saves the data to a .json file. Convert this data into a list that contain the data for each element. At this point create a dictionary of hydrogen with the keys being the column densities, and the values being the redshift. Sort this dictionary by column density. Next create a dictionaries of the systems with column density greater than 14 with redshifts mapping to elements found there, of all hydrogen systems, and all elements found each mapped to a redshift for later reference. The second data set is acquired from the SDSS DR7 website SQL query, and is a list of all galaxies and stars within a 0.5-degree window of the analyzed object. This data set contains coordinates in degrees, object IDs, redshifts and redshift uncertainties. Store this data in a list of dictionaries with the keys being an object's ID, redshift, redshift error, and coordinates, each mapping to the respective quantity. Next, create a dictionary of the strong hydrogen systems, and map them to a list of potential galaxies the have redshifts with uncertainties that contain the strong hydrogen. From here we need to filter these lists by calculating the distance from each object to the line of sight, and just keeping the host galaxy with the minimum distance. The distance is calculated by using cosmology; I use the redshift to find the actual distance, the coordinates to find the angular distance between in arc seconds, and then create a scale factor because the galaxy gets smaller the further away one looks due to expansion. By multiplying the angular distance by the scale we get a distance in kilo parsecs (kpc). Display the result of the gas location, the galaxy location (with uncertainties), the size of the CGM, and finally a list of the metals found in the Galaxy. Finally plot the systems of gas, and the objects corresponding to the system, and print the lower bounds of the CGM for each system.

Results:

1. I was able to map each system of strong Hydrogen to a galaxy, all with good fits except for one. The bad fit was due to a very large uncertainty in the distance of the host galaxy. The conclusion that I calculated is that CGM extends anywhere from about 0 to 850 kpc. My program prints out all of the distance of the gas, distance and error of the galaxy, the galaxy ID, the lower estimate for the CGM size, and the elements found in the CGM. This falls in line with the values cited in Sowgat Muzahid's paper on the CGM from 2014 (1).



Part of output from program:

Gas Location (Redshift): 0.26378

Galaxy Location (Redshift): 0.271857 \pm 0.046805

Galaxy ID (SDSS Catalogue): 587732134303629397

Galaxy Halo (kpc): 46.1671835718

Elements Found: ['CIII', 'CIII', 'HI', 'NIII', 'OVI', 'SiIII']

2. What can the compositions of these CGM tell us about the host galaxy? The more metallic a galaxy, the older a galaxy is due to the fact that elements heavier than helium are forged in dying stars and supernovae, both of which take time. The older Galaxies in this analysis are the systems found around redshift 0.26. This system contains carbon, nitrogen, oxygen, and silicon. This galaxy is the best example of the program because the galaxy is quite visible and close to the line of sight providing excellent data. In addition, using the ID of the Galaxy, the Galaxy image and spectrum can be looked up on the SDSS database to verify the composition of CGM. As more and more spectrum get analyzed, the more data points of the lower limits of the CGM size, and in the future we can more accurately map these out. Future surveys to keep an eye on will be from the Large Synoptic Survey Telescope and James Web Space Telescope.

Reproducing Results:

Reproducing results should be simple. Obtain an analyzed QSO spectrum as a .json file, and a dataset of galaxies around the QSO by using the SQL query at the SDSS website. A sample SQL input can be found in the Dataset section of this report. Using these, load the files into my program, and the system will be plotted along with all of the strong Hydrogen being mapped to their host galaxies.

Example code from command line would be (Commands in bold):

-Make sure the program is in the same directory as the json file, and the galaxy text file.

Python Final_project.py

-At this point you will be prompted to input the file names.

Ex:

Spectrum json:

dusty3_J1016+4706.json

Galaxies data:

J1016_Galaxies.txt

This will result in a plot being generated along with the distances of gas systems, galaxies (with uncertainties), galaxy IDs, CGM size, and a list of metals.

Work Plan Evaluation:

- Organize Data from Spectrum
 - Scrub and Organize
 - Time: 3 hours
 - Actual Time: 3 hours
- Access Data from SDSS DR7
 - Organize Data to be usable
 - Time: 3 hours
 - Actual Time: 5 hours
- Make Calculations of distances
 - Angular Distance
 - Time: 1 hour
 - Actual Time: 1 hour
 - Actual Distance
 - Time: 1 hour
 - Actual Time: 1 hour
- Plot Results
 - Time: 1 hour
 - Actual Time: 1 hour
- Print Results

Overall, my work plan was fairly accurate. I did run into some troubles making the program more accessible with user input from the command line.

Testing:

I tested my code by using assert statements throughout the program. These assert statements test the length of data structures, and the values generated. To ensure that these were accurate, I did calculations on the side, and then tested the generated output vs. my output. These results should be trusted because all of the assert statements passed, and in addition I have verified a couple of systems' host galaxies with my research professor.

Live Presentation or Video:

Live Presentation in class.

Collaboration:

-I used one section of code from my research group to take the json file, and put the data into a list. This section of code is called spectrum_to_lst in my program, and in the string doc it is said that I did not write it. Other than this, I worked by myself.

Bibliography:

1. Muzahid, Sowgat. "PROBING THE LARGE AND MASSIVE CIRCUMGALACTIC MEDIUM OF A GALAXY AT $z \sim 0.2$ USING A PAIR OF QUASARS." *The Astrophysical Journal* 784, no. 1 (2014): 5. doi:10.1088/0004-637x/784/1/5.