Applying Hierarchical Clustering to Broad Absorption Line Profiles for Quasar Classification

Nathalie Thibert

Collaborators:
Dr. Sarah Gallagher
Dr. Mark Daley
Dr. Aycha Tammour

Image credit: NASA
Broad Absorption Line (BAL) Quasars

Radiatively-driven Outflow
Supermassive Black Hole
UV Continuum

Edge-on Accretion Disk

- $10^{18}$ cm
- $10^{17}$ cm
- $10^{16}$ cm
- $10^{15}$ cm
Diversity of C IV BALs

Why are these profiles different?
BAL Quasar Catalog & Data Reduction

- **Sample:** 1,084 BAL quasars \((1.68 \leq z \leq 4.93)\) from SDSS DR5 (Gibson et al., 2009)
- **Method:**
  1. fit local continuum and prominent, broad emission lines \((1320-1800 \, \text{Å})\)
  2. normalize and “compare” spectra \((1400-1550 \, \text{Å})\)
BAL Quasar Catalog & Data Reduction

- Sample: 1,084 BAL quasars (1.68 ≤ z ≤ 4.93) from SDSS DR5 (Gibson et al., 2009)
- Method: (1) fit local continuum and prominent, broad emission lines (1320-1800 Å) 
  (2) normalize and “compare” spectra (1400-1550 Å)
Pearson Product-moment Coefficients

Strong Positive Correlation:

Weak/No Correlation:

Strong Negative Correlation:

$\rho = +0.948$

$r = 4.14E-08$

$r = -0.853$
Matrix of Correlation Coefficients
Unclustered Distance Matrix
Results from Hierarchical Clustering

scipy.cluster.hierarchy.linkage
scipy.cluster.hierarchy.dendrogram
Results from Hierarchical Clustering

Identify ~10 clusters

scipy.cluster.hierarchy.linkage
scipy.cluster.hierarchy.dendrogram
Dendrogram of the Merging Process
Example: Broad, High-velocity C IV Troughs

Broad C IV troughs that extend to high maximum outflow velocities
Broad, High-velocity C IV Troughs

High outflow velocities

A

B1

B2

Normalized Flux Density [10^{-17} erg s^{-1} cm^{-2} \AA^{-1}]

Rest Frame Wavelength [Å]

May 3rd, 2016

Nathalie Thibert

Great Lakes Quasar Symposium 2016
Broad, High-velocity C IV Troughs

Emission line ratios as a proxy for the hardness of the SED.

Al III/ C III] high → softer SED

IP(Al III) = 18.8 eV
IP(C III) = 24.4 eV
Conclusions/Next Steps

- We apply hierarchical clustering to the C IV troughs of 1,084 BAL quasar spectra. We see at least 10 distinct clusters.

- Certain emission and absorption line properties are seen to group together in distinct categories, which could imply a common cause.
Additional Slides
Example: Shallow, Low-velocity C IV Troughs

Shallow and extending to low maximum outflow velocities
Shallow, Low-velocity C IV Troughs

Gradual initial onset of velocity

High-velocity component?
Shallow, Low-velocity C IV Troughs

Emission line ratios as a proxy for the hardness of the SED.

Al III/ C III] low → harder SED

$\text{IP(Al III)} = 18.8 \text{ eV}$

$\text{IP(C III)} = 24.4 \text{ eV}$
Example: Deep, Any-velocity C IV Troughs

Deep and extending to any outflow velocity
Deep, Any-velocity C IV Troughs

Gradual initial onset of velocity

He II + O III] peaks higher → harder SED

IP(He II) = 24.6 eV
IP(O III) = 35.1 eV
Deep, Any-velocity C IV Troughs

Emission line ratios as a proxy for the hardness of the SED.

\[
\text{IP(Al III)} = 18.8 \text{ eV} \\
\text{IP(C III)} = 24.4 \text{ eV}
\]

Al III/C III low $\rightarrow$ harder SED

Softer SED allows wind to reach higher outflow velocities
Pearson Correlation Coefficients

- Measures how two sets of data are correlated with one another.

\[ r = \frac{\sum_{i=1}^{n} ((x_i - \bar{x})(y_i - \bar{y}))}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \]

\[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \]

- Simple to compute in Python:

```python
>>> numpy.corrcoef( ... )
```

Hierarchical Clustering Algorithm

Bovermann et al. (2008)

Generalize to 1084-dimensional space...

Spectrum 1 feature vector: \((0.0, d_{1,2}', d_{1,3}', \ldots, d_{1,1084})\)