Use of machine learning techniques to investigate trends in AGN/Quasar activity and evolution

Rohin Sant (University of Arizona - rsant@arizona.edu) & **Stéphanie Juneau** (NSF NOIRLab)

Abstract

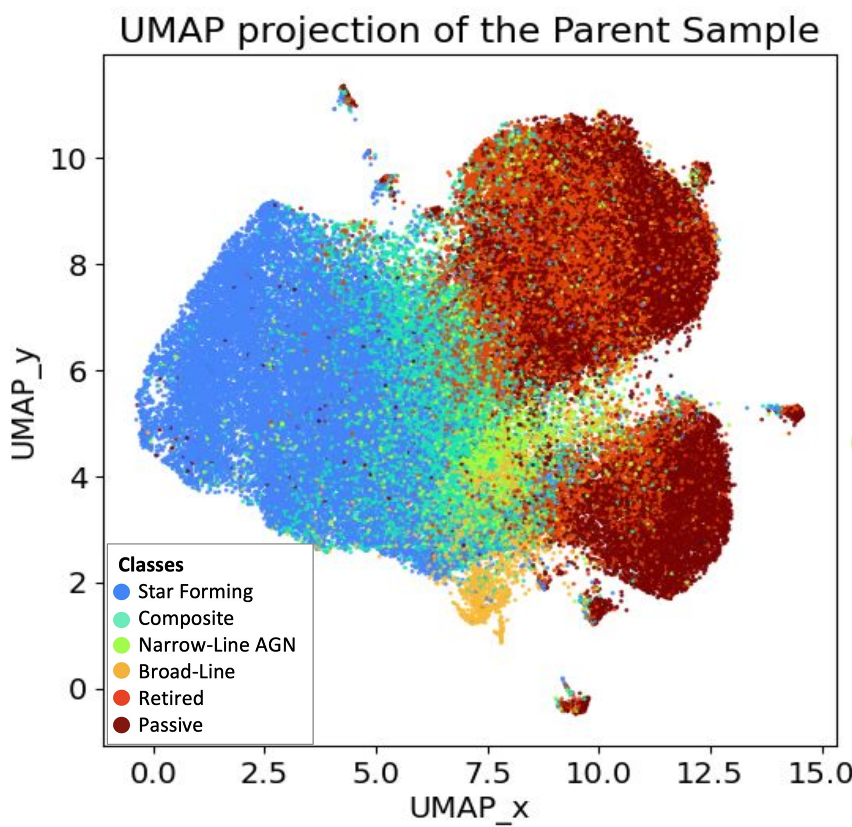
We investigate how a machine learning algorithm, Uniform Manifold Approximation and Projection (UMAP), gives insight about active galactic nuclei (AGN) and quasar activity of over 200,000 low-redshift galaxies with spectra from the Sloan Digital Sky Survey. Using traditional broad-line and narrow-line diagnostics on the sample, we find that AGN (broad line and narrow line) and quasars occupy distinct regions on the UMAP space. We investigate trends with AGN luminosity and create a new UMAP projection for the AGN and quasars alone to search for trends more closely related to black hole growth. This work demonstrates the potential of unsupervised ML tools to reveal physically meaningful patterns in large astronomical datasets, offering new ways to study black hole growth and AGN diversity.

Background

- Active Galactic Nuclei (AGN) are powered by black holes, emit energetic radiation and can potentially influence their host galaxies:
  - *What can we learn about AGN and galaxies from their spectra?*
- Spectra of galaxies encode a lot of information on AGN and galaxies but it is challenging to interpret these large multi-dimensional datasets
- UMAP (McIness et al. 2018) is a useful technique to help visualize high dimensional data, such as large spectroscopic surveys (SDSS, DESI):
  - Unsupervised ML algorithm that deals with dimensional reduction.
  - Calculates the likelihood of the neighboring data points being connected.
- We seek how UMAP gives insight into AGN/Quasar activity of galaxies beyond traditional AGN and galaxy classification.

Previous Findings

- Pat et al. (2020) used unsupervised ML to study >200,000 low-redshift galaxies from the Sloan Digital Sky Survey (SDSS)
  - A Probabilistic Auto-Encoder (Böhm et al. 2023) reduced spectral data to 10D; then UMAP reduced to 2D for better visualization.
- Galaxies color-coded with standard spectral classification **(Fig. 1)**



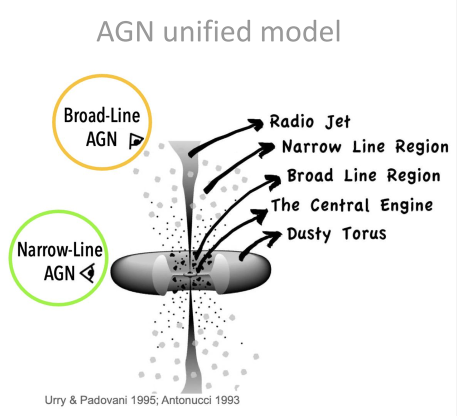
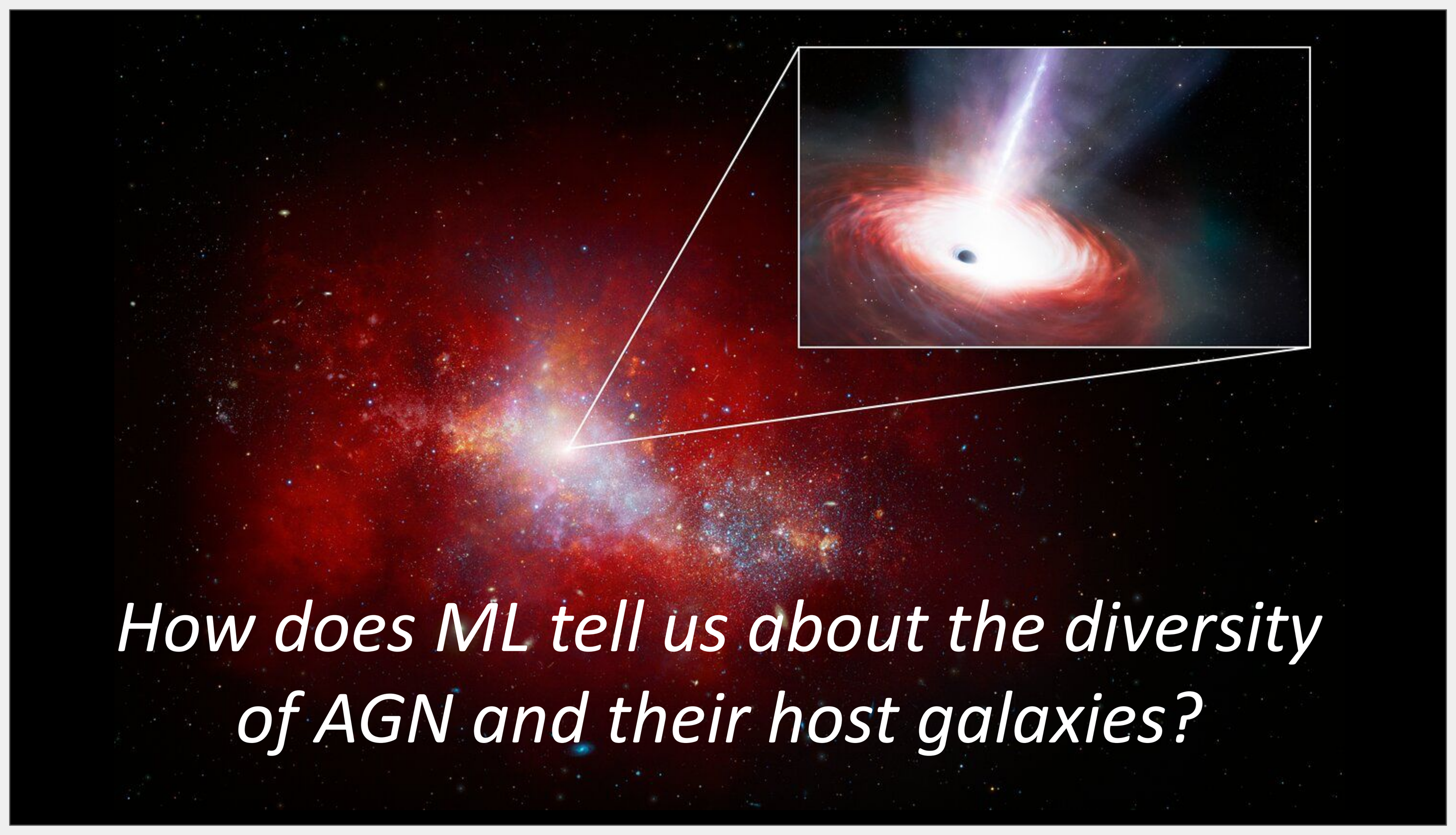


Fig. 1: UMAP Diagram (Classification of Galaxies): Despite the classification labels not being used in the ML training, we see intriguing trends in galaxies occupying distinct regions for different galaxy classes.

This Work: Focusing on AGN

- **Parent sample:** training set from Pat et al. 2020 comprising 209,462 low-redshift SDSS galaxies (0.05<z<0.36)
- Galaxies classified based on traditional emission line diagnostic diagrams (BPT, WHAN) into six categories **(Fig. 1)**
- ***AGN and quasar subsample:*** 40,628 composites, NL-AGN, BL-AGN
- Value-added catalogs crossmatched with our samples:
  - SDSS DR12 Portsmouth Emission-Lines (Thomas et al. 2013): emission line fluxes and their errors [except it lacks some BL-AGN]
  - DESI DR1 AGN/QSO (Juneau et al. in prep.): AGN classification and line fluxes from FastSpecFit (Moustakas et al.) with narrow/broad components



Methods

- We compute line ratios from both SDSS and DESI (Fig. 2) and the AGN luminosity from [OIII]5007 (Lamastra et al. 2023)
- We clean the broad-line samples by applying a cut to the Hα equivalent width [EW(Hα)> 3Å], which removes wrongly labeled galaxies previously reported by Pat et al. 2020
- We apply UMAP to the 10D latent space from Pat+2020 and recreate the 2D projection (Fig. 1)
- We apply UMAP on just the AGN/Quasar subsample (Fig. 4) with a study of the hyperparameters (`n_neighbors, min_dist`), finding that a low number of neighbours and low minimum distance focuses more in local structure of the data while the vice versa focuses on the global structure.

AGN Classification in SDSS vs. DESI

We compare BPT diagram classification between SDSS and DESI:

- **Most cases are consistent**, except: (1) Some Narrow-line AGN are displaced in the Star-Forming region; (2) Broad-Line AGN have lower [NII]/Hα

Future work: are differences due to line fitting methods and/or fiber aperture size?

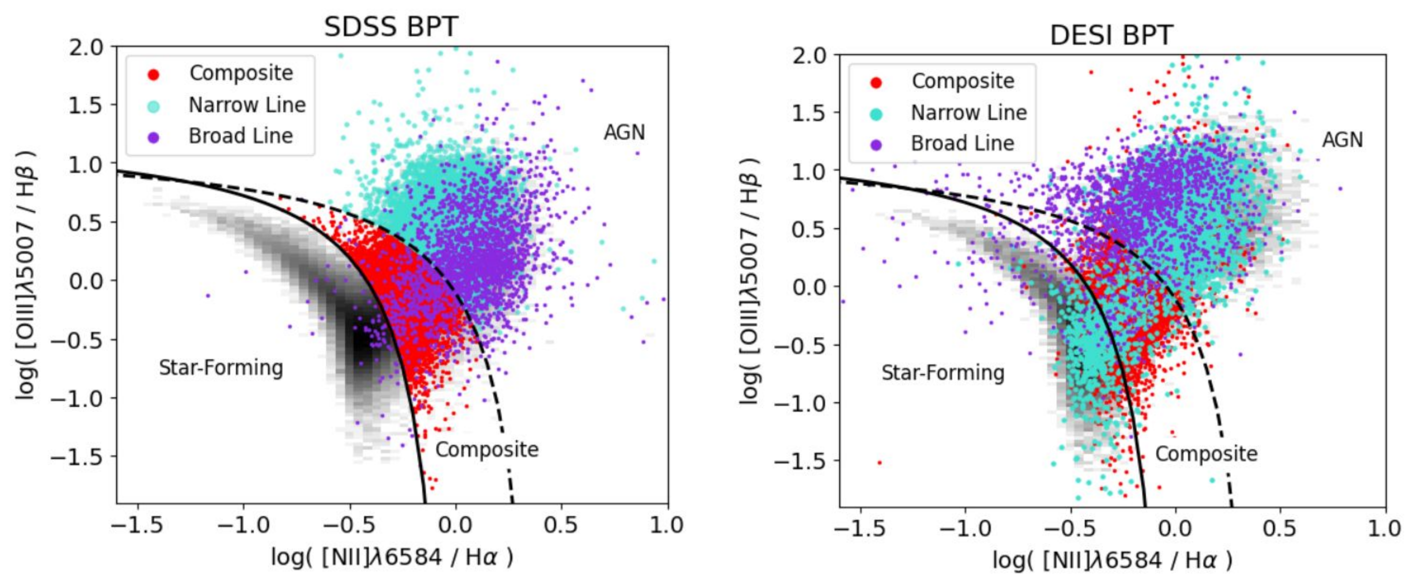


Fig. 2: BPT Diagram for SDSS (left) **and DESI** (right) **spectral measurements.** Some narrow lines in DESI BPT spillover into the Star Forming Region.

Trends with ionized gas properties

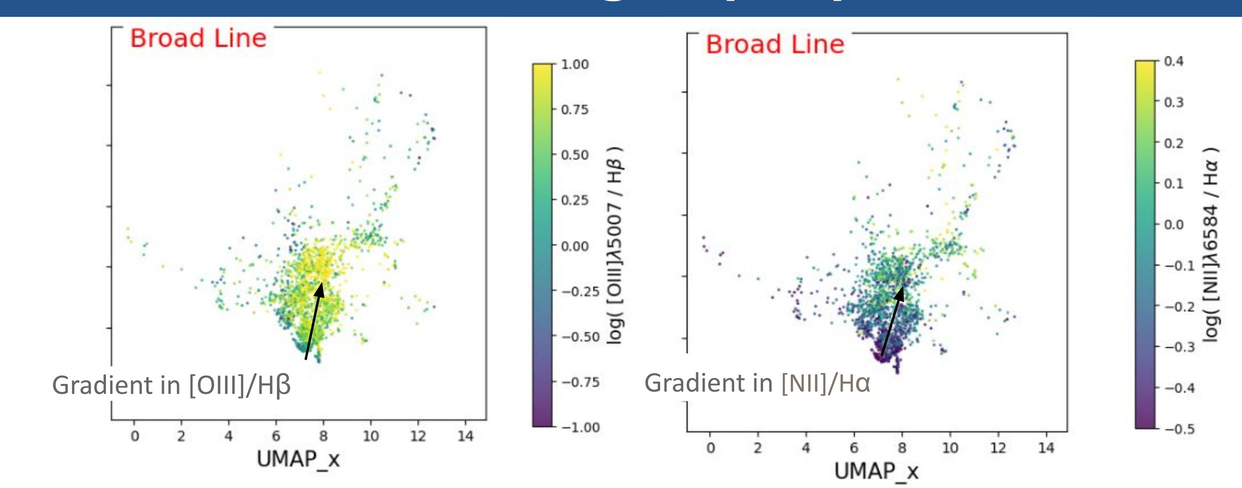


Fig. 3: Broad-Line AGN color-coded by [OIII]/Hβ **and** [NII]/Hα. We see clear gradients in both line ratios, reaching comparatively low values toward the bottom part. This is not what we expect for energetic quasars, which might have high ionization parameters (high [OIII]/Hβ) and metal-rich gas (high [NII]/Hα). Further study is required to understand why these ratios appear lower than expected.

Comparison of UMAP projections

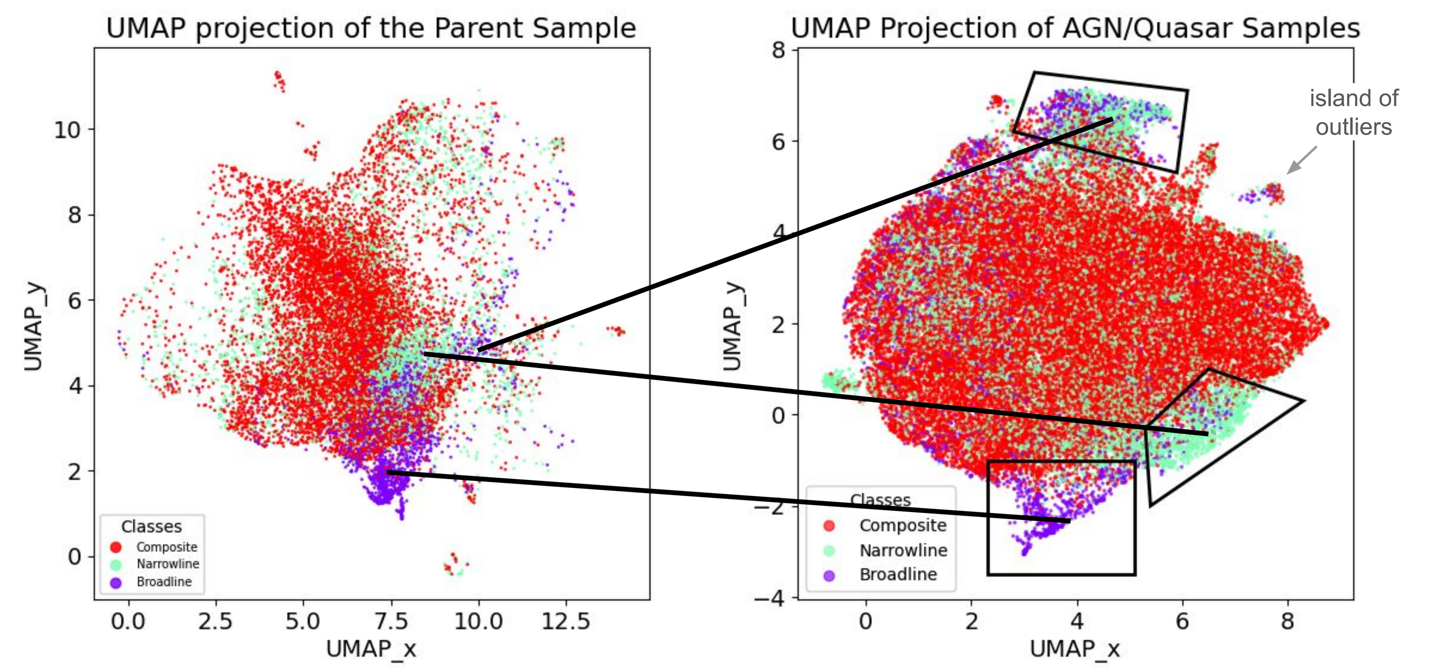


Fig. 4. UMAP Diagram (Classification of AGN/Quasars): We compare trends between composite, narrow-line and broad-line samples on the parent sample (left) and new AGN only (right) UMAP. We mark new regions of interest (black polygons) and point to their location on the original UMAP. Broad-lines mostly form a distinct region near the bottom, with narrow-lines forming near the bottom right and mixed with some broadlines at the top. Composites tend to overlap with the other samples.

Future Work: Look at the spectra in regions of interests to get insights on trends.

Trends with AGN luminosity

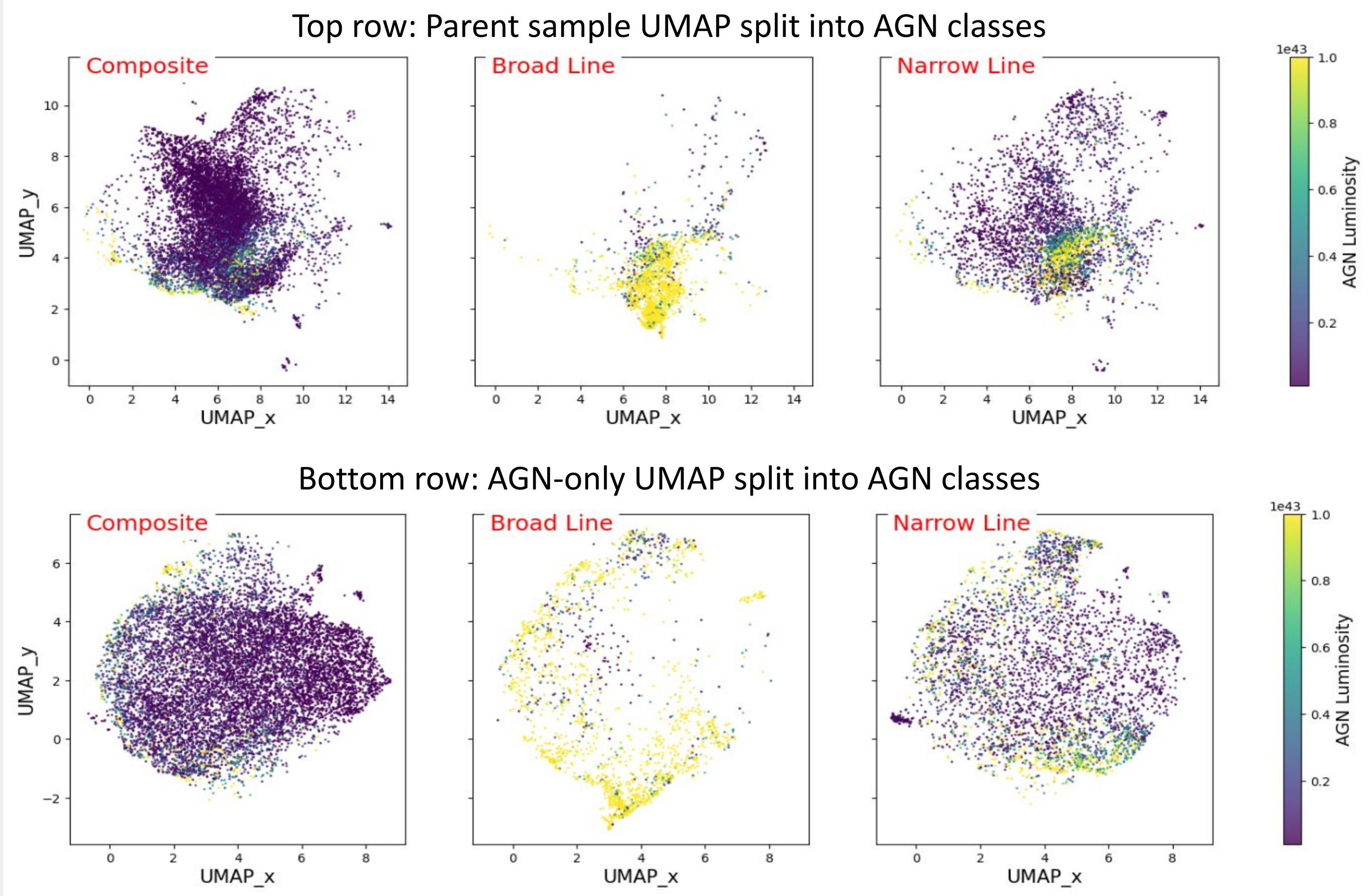


Fig. 5. Regions of UMAP color coded by AGN Luminosity (Lamastra et al. 2023): Most AGN luminous galaxies present in broad-line AGN, which include quasars. This is expected as quasars accrete a lot of material around supermassive black holes. Narrow-line and composite samples are overall less luminous, except in regions near broad-line samples. In details, there are differences between the two UMAP projections (e.g., distinct AGN-luminosity gradient in narrow-lines in top).

Conclusions & Future Work

- UMAP classifies the AGN/Quasar samples into regions such as Broad-line, Narrow-line, mixed regions overlapping with Composite samples.
- Both UMAP projections reveal patterns between AGN luminosity – a proxy for black hole growth – and the type of AGN.
- **Future:** Investigate the "islands" of outliers using spectral analysis.
- **Future:** Investigate the discrepancy between the SDSS & DESI BPT diagram.

References

- Pat, Juneau, Böhm et al. (2020), CURASS. ASP Conference Series, Vol. 525. ASP, 2022, p.67
- McInnes, Melville, Healy et al. (2018), arXiv preprint arXiv:1802.03426.
- Lamastra, Bianchi, Matt et al. (2009), A&A, Volume 504, Issue 1, 2009, pp.73-79.
- Böhm, Kim, Juneau et al. (2023), MNRAS, Volume 526, Issue 2, December 2023, Pages 3072–3087
- Thomas, Steele, Maraston et al. (2013), MNRAS, Volume 431, Issue 2, p.1383-1397
- Moustakas, Buhler, Scholte et al. (2023), Astrophysics Source Code Library, record ascl:2308.005

Acknowledgments: This work was supported by the NSF-Simons CosmicAI institute, the University of Arizona's TIMESTEP program and made use of data and services from the NSF NOIRLab Astro Data Lab. This work is based in part on data from the Sloan Digital Sky Survey (SDSS) and the Department of Energy Dark Energy Spectroscopic Instrument (DESI).