

Active Dwarf Galaxy Database II: Connections between Host Galaxy Properties and Black Hole Accretion Signatures

ERIK J. WASLESKE ^{1,2} VIVIANNE F. BALDASSARE ² AND CHRISTOPHER M. CARROLL ²

¹*Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 West Green Street, Urbana, IL 61801, USA*

²*Department of Physics and Astronomy, Washington State University, Pullman, WA 99163, USA*

(Received 18 June 2025; Revised 1 October 2025; Accepted 16 October 2025)

Submitted to ApJ

ABSTRACT

We investigate the connection between accretion signatures and host galaxy properties in the context of how active dwarf galaxies are identified. We use the database constructed in Wasleske & Baldassare (2024) which contains dwarf galaxies that were selected as active galaxies by optical spectroscopy, infrared colors, X-ray brightness, and photometric variability. Multi-wavelength archival data was used to consistently apply all of these methods to every galaxy within this compiled set. The cross application of these methods resulted in a diversity of sub-populations identified as active by some set of these techniques. In this paper, we estimate host galaxy properties from spectral energy distribution models. We connect the active galactic nuclei (AGN) signatures to our estimated host galaxies’ properties using statistical dimensionality reduction methods. We find that dwarf AGN selected by infrared colors are the most distinct population, with the highest star formation rates and lowest stellar masses. We also find some other key population differences, such as the broad line AGN having significantly higher AGN luminosities. X-ray and variability selected AGN have higher average star formation rates than those selected with optical narrow line spectroscopic diagrams. Our connections to the host galaxy parameters potentially point to the sub-populations representing different epochs of the evolution of accretion.

1. INTRODUCTION

The correlation between the mass of host galaxies and their central massive black hole (MBH) has been well studied (Ferrarese & Merritt 2000; Gebhardt et al. 2000; Kormendy & Ho 2013). These MBHs can affect host galaxy properties such as morphology and star-formation (SF; Cattaneo et al. 2009). Conversely, the gas, dust, and SF within the galaxy can mask the emission from the active MBH. This adds to the difficulty of compiling a complete census of active galaxies.

There has been extensive work done to establish a set of methods to identify active MBHs in the center of galaxies (e.g., Ho et al. 1997; Ulrich et al. 1997; Berk et al. 2004; Just et al. 2007; Desroches et al. 2009; Assef et al. 2010; Stern et al. 2012). Each method selects MBHs based on different emission signatures stemming from the accretion of material within the active galactic

nucleus (AGN). The canonical picture of an AGN comprises the central accreting black hole (BH) surrounded by an accretion disk, with an X-ray corona, enveloped by a dusty torus with polar jets (Antonucci 1993). With each type of observation, different selection biases are introduced from limitations in observing the radiation from these components.

Spectroscopic methods (e.g., Baldwin et al. 1981; Kewley et al. 2006; Shirazi & Brinchmann 2012) use narrow emission line ratios to separate signatures of accreting BHs from stellar processes. These diagrams can also distinguish Seyfert galaxies from low-ionization emission line ratio (LINER) type galaxies. For dwarf galaxies ($M_* \leq 10^{9.5} M_\odot$) which contain less massive central BHs compared to more massive galaxies, their lower accretion power may not outshine the stellar processes. Additionally, the lower host galaxy metallicities alter the value of the line ratios to move active galaxies into the star forming region (Groves et al. 2006; Reines et al. 2013). This increases the potential for active BHs in dwarf galaxies

to be misclassified. The models of Cann et al. (2019) showed that with decreased mass, a galaxy shifts out of the AGN region of the standard BPT diagram (Baldwin et al. 1981) as the lower mass BHs have more extended partially ionized zones. Broad emission lines—a signature of Doppler broadening due to the close proximity of gas to the central BH—are less frequently observed in dwarf candidates (Reines et al. 2013) due to lower BH masses.

Bright X-ray emission has long been an observed feature of AGN (Brandt & Hasinger 2005; Greene & Ho 2007; Desroches et al. 2009; Lemons et al. 2015; Plotkin et al. 2016; Birchall et al. 2020). The strength of this X-ray emission is known to scale with BH mass (Merloni et al. 2003). Although X-rays are less susceptible to the effects of obscuration, one must rule out other possible sources of X-ray emission. X-ray binaries and supernovae remnants serve as a sources of X-ray emission, leading to a tight correlation of hard X-rays to star formation rate in star forming galaxies (Ranalli, P. et al. 2003). In turn, these phenomena must be ruled out to determine if the X-ray emission is from accretion. This source of contamination is more prevalent when searching for weaker X-ray emissions of active BHs in dwarf galaxies which can be masked by enhanced emission from low metallicity stellar populations (Lehmer et al. 2021).

Infrared (IR) colors measured from the Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010) have been used to identify AGN (Jarrett et al. 2011; Stern et al. 2012; Assef et al. 2018; Hviding et al. 2022) as the WISE wavelengths are sensitive to the obscuring torus of the AGN. These WISE IR colors are biased towards AGN that dominate the host galaxy stellar emission. However, these selection techniques are heavily influenced by dust and star formation, as extreme starburst activity within galaxies can mimic AGN within these selections (Satyapal et al. 2018). Hainline et al. (2016) found the majority of optically selected dwarf galaxies, with active signatures from their narrow emission lines, are not selected using these IR color criteria.

The intrinsic variability of AGN has been used as a tool to identify these objects within time domain data with good success (e.g Schmidt et al. 2010; MacLeod et al. 2011; Baldassare et al. 2018; Burke et al. 2022) and is of particular interest as new time domain surveys come online. Recent works have searched for correlations of variability to characteristics of the AGN (eg. characteristic dampening time to black hole mass Burke et al. 2021; Wang et al. 2023) However, technical survey details can limit the detection of variability. The baseline of observations, cadence, and gaps in data can

impact the modeling of the varying stochastic signal (Kozłowski 2017). Additionally, as variability is investigated often from optical imaging, handling host galaxy contribution requires great care to accurately characterize variations from the AGN’s emission, especially if it is of lower power.

The inherent issues of each technique listed above lead to an incomplete understanding of the population of active dwarf galaxies. To collect the most complete and correct sample of active dwarf galaxies, one must implement multi-wavelength observations. Accurate populations can be used to constrain seeding mechanisms of central galactic BHs and the evolutionary relationships with their host galaxy (Greene et al. 2020 and references therein).

In Wasleske & Baldassare (2024) we compiled a set of 733 active dwarf galaxies that were initially identified using optical spectroscopy, X-ray, infrared (IR), and optical photometric variability. We collect available archival data and applied a uniform set of selection techniques to this database. After the application of methods using optical line diagnostic diagrams, X-ray photometry, IR colors, and photometric variability, identifying a subsample of galaxies selected as AGN from each method. We then cross applied these methods to these subsets to investigate the overlapping populations identified by these methods. This resulted in a lacking agreement between these subsets.

In this paper, we investigate the biases in these selections and search for a parameter space that distinguishes these populations from another. We first estimate host galaxy properties and connect them to selection methods. We model the spectral energy distributions (SEDs) to investigate the distribution of host parameters within the AGN subsamples. We then use t-distributed stochastic neighbor embedding (t-SNE, van der Maaten & Hinton 2008) as a clustering tool to group galaxies based on a combination of the host galaxy parameters, which in turn allows us to decouple galaxies identified with multiple AGN signatures.

This paper is organized as follows: Section 2 we compile additional photometry for our analysis; Section 3 describes our construction of spectral energy distribution (SED) models using this photometry and our application of t-SNE clustering; Section 4 describes the results drawn from the selection subsamples and the t-SNE clusters; Section 5 considers the implications of these results and the influence host galaxy properties have on the selection of active dwarf galaxies. We assume a Λ CDM cosmology with parameters $h = 0.7$, $\Omega_m = 0.3$ and $\Omega_\lambda = 0.7$ (Spergel et al. 2007).

2. DATA

We start with the resulting active dwarf galaxy database (ADGD) from (Wasleske & Baldassare 2024; hereafter W24), using these positions to collect more accurate photometry and results where needed.

Our sample is comprised of the resulting 702 active dwarf galaxies identified from a uniform application of AGN-selection techniques. To accurately model and constrain the SEDs of our sample, we compiled additional photometry to supplement W24. This involves collecting optical photometry as no optical photometry was used in W24, and additional IR and UV data.

To collect SDSS *ugriz* observations from DR 17 (Abdurro'uf et al. 2022), we use `astroquery` to search for matches within a $3''$ radius. We use the `cModelMag` values for these five bands of the closest source, restricting to only clean photometry (ie. only keep photometry values for whose 'CLEAN' value is 1). Extinctions of these magnitudes were used to correct `cModelMag` to calculate the dereddened magnitude in each band. No values were found for 46 of our objects.

We use the position of the SDSS source to query the rest of the photometry discussed here. For the 46 objects with no SDSS photometry, we used their positions from the ADGD to query the rest of the photometry listed below. Of the 46 objects with no SDSS photometry, 20 have photometric values in at least one band of Pan-STARRS DR2 (Chambers et al. 2016; Flewelling et al. 2020), ten of which have values in each *grizy* filter. For consistency, we opt to not include this set of Pan-STARRS photometry into our analysis.

Furthermore, we collect photometry from the Wide-Field Infrared Survey Explorer (WISE, Wright et al. 2010) AllWISE Source Catalog (Wright et al. 2019) for each galaxy, within a $3''$ radius, across the four infrared bands to supersede the use of the WISE All-Sky Release catalog in W24. We compile the profile-fitting photometry magnitudes for these four bands (W1, W2, W3, W4 bands with wavelengths 3.4, 4.6, 12 and $22\ \mu\text{m}$ respectively). W1, W2, and W4 data was available for 694 of these objects, with W3 data available for all but one of these 694 objects. For UKIDSS/UKIRT (Lawrence et al. 2007), we use `astroquery` to first search the Large Area Survey (LAS) in a Y, H, J and K bands. We collected Petrosian magnitudes for each of these bands from the UKIDSSDR11PLUS catalog search within $3''$ of the galaxy positions. This data was available for 302 objects within the ADGD. Petrosian magnitudes should have good agreement with the `cModelMag` of SDSS (Strauss et al. 2002).

Additionally, we searched for available Galaxy Evolution Explorer (GALEX) near($1750 - 2800\ \text{\AA}$) and

far($1350-1750\ \text{\AA}$) ultra-violet fluxes within $3''$ of each position. We used `astroquery` to search the GUVcatAIS GR 6+7 catalog (SourceRevised catalog of GALEX UV sources (GUVcat_AIS GR6+7), Bianchi et al. 2017). NUV values was available for 504 of the objects and FUV values for 442 of the objects.

The X-ray photometry is taken directly from the ADGD as found by the Chandra X-ray Observatory and XMM-Newton analysis in W24. We compiled all of these values as inputs for our spectral energy distribution modeling described below.

3. METHODS

We created SED models to estimate the host galaxy properties. Ramos Padilla et al. (2021) used SED models in unison with machine learning algorithms to test to standard AGN model and its assumption of classification based on viewing angle, showing the utility of this combination. In the following, we first discuss the construction of our SED models by illustrating the modules and parameter grids used. After, we list the machine learning algorithms implemented to draw the AGN classification-host galaxy parameter connection.

3.1. Spectral Energy Distribution Modeling

We used X-CIGALE V2022.0 (Yang et al. 2022) to model the SEDs of the galaxies in our sample. CIGALE (Noll et al. 2009; Boquien et al. 2019) is a Python-based code created to efficiently model the FUV-to-radio spectrum of galaxies. It estimates galaxy properties by using multi-wavelength photometric observations of galaxies to constrain models of stellar, dust attenuation, nebular and AGN emission. CIGALE uses energy balance principles to give results for large samples of galaxies within much shorter computational time than solving the 3-dimensional radiative transfer equation whilst still having high accuracy on parameter estimates. The properties estimated from these SEDs use Bayesian approach by weighing all models by their goodness-of-fit. These methods includes observational uncertainties and the degeneracy between the parameters. X-CIGALE builds upon the CIGALE by including X-ray photometry fitting and a model of polar dust in AGNs, making this fitting code ideal for AGN.

We present the modules and parameters used in our implementation of X-CIGALE to generate SED models of the galaxies in the ADGD. We base our parameter grid off of the one used in Wasleske & Baldassare (2023), with modifications listed below and outlined in Table 1.

First we include a delayed star formation history module (`sfhdelayed`) to adequately account for early- and

X-CIGALE Parameters

Parameter	Value	Description
Star Formation History: sfhdelayed		
τ_{main}	100, 500, 1000, 5000 Myrs	e-folding time for main stellar population model.
age_{main}	1000, 2000, 5000 Myrs	Age of the main stellar population in the galaxy.
Single-age Stellar Population: bc03 (Bruzual & Charlot 2003)		
IMF	1	Chabrier (2003) initial mass function
metallicities	0.0001, 0.0004, 0.004, 0.008, 0.02	Range for stellar population
Separation Age	10 Myrs	Differentiate old & young stellar populations
Nebular Model: nebular		
$\log U$	-2.0	Ionization parameter.
z_{gas}	0.004, 0.008, 0.014, 0.033	Range for gas metallicity.
Dust Attenuation: dustatt_modified_starburst (Modified Calzetti et al. 2000 Attenuation Law)		
Extinction law	MW, LMC	For attenuating the emission line flux.
Dust Emission: dl2014 (Draine et al. 2013)		
Q_{PAH}	2.5	Mass fraction of PAH.
U_{min}	1.0	Minimum radiation field.
α	2.0	Powerlaw slope dU/dM propto U^α
γ	0.1	Fraction illuminated from U_{min} to U_{max} .
AGN model: skirtor2016 (Stalevski et al. 2016)		
frac_{AGN}	0.0, 0.1, 0.25, 0.5, 0.6, 0.7, 0.8, 0.9	Fraction of AGN torus to the IR total luminosity.
E(B-V)	0.0, 0.025, 0.05, 0.075, 0.10, 0.125, 0.15, 0.175	Excess Color in polar direction.
i	10, 30, 50, 70°	viewing angle with respect to the AGN.
X-ray (from skirtor2016 AGN)		
α_{OX}	-2.0, -1.75, -1.5, -1.25, -1.0	Power law slope defined in Just et al. (2007).
Γ	1.8	Photon index of the intrinsic AGN x-ray spectrum

Table 1. Grid parameters for the X-CIGALE SED modeling.

late-type galaxies (Boquien et al. 2019) with e-folding times of 100, 500, 1000, or 5000 Myr. Moreover, the stellar model of Bruzual & Charlot (2003) is suited to fit stellar populations with ranging metallicities and ages. We allow for fits with metallicities of 0.0001 – 0.02 for a Chabrier initial mass function (Chabrier 2003). We include the modified dust-attenuated starburst module from Calzetti et al. (2000).

For our AGN contribution, we used the SKIRTOR (Stalevski et al. 2012; Stalevski 2012; Stalevski et al. 2016) module. This module implements a two-phase dusty torus of low density medium filled with high density clumps. We set the inclination angle to the set of 10°, 30°, 50°, or 70° and the AGN fraction to values from 0.0 to 0.9.

The dust contribution is modeled by dl2014. This module is based of the dust model of M31 (Draine & Li 2007), which assume dust comprised of carbonaceous grains and amorphous silicate grains and variable polycyclic aromatic hydrocarbon abundances. The size distribution of the grains is consistent with the extinction

and IR emission of diffuse interstellar medium within the solar neighborhood.

We only kept those objects whose SED model has $0.5 < \chi_r^2 \leq 10$. This left 461 galaxies to analyze. The number of these galaxies within each sub-sample is given in the first column of Table 2. The preliminary values for host galaxy properties that can be derived from our data (i.e. the SFR derived in W24 using corrected H α measured from spectrum) were not used to inform the priors for our SED modeling. As both values taken from derived relations the X-CIGALE modeling have systematic uncertainty, we chose to keep our SED models agnostic to previous results.

From these models we collect Bayesian estimates of host galaxy dust and stellar luminosity, the stellar metallicities, and the AGN luminosity. Example models are shown in Figure 1. We will be using redshift and host galaxy stellar mass from W24 alongside the estimated values of stellar metallicity, L_{dust} , L_{AGN} , L_{stellar} and E(B-V) for the input to all machine learning algorithms explored below. L_{AGN} is sum of the AGN disk and dust re-emitted luminosity taken from the SKIRTOR

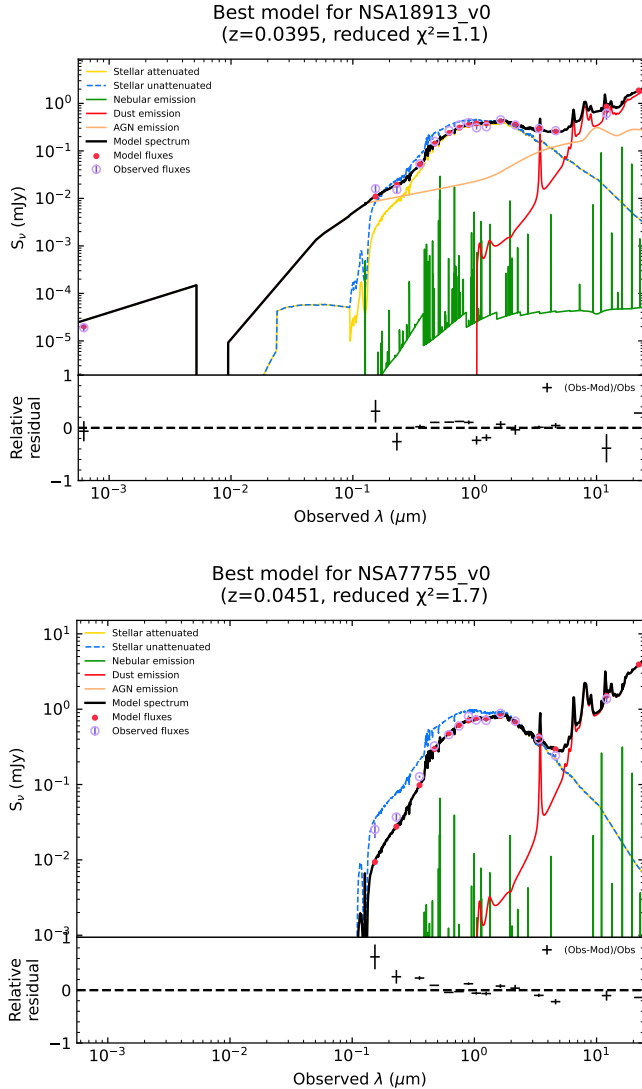


Figure 1. SED models for NSA 18993(top) and NSA 77755(bottom). NSA 18993 was selected as an AGN via the BPT and [He II] diagrams, as a Seyfert on the [S II] diagram, and from its broad H α and X-ray emissions. In contrast, NSA 77755 was only selected based on optical variability.

module. L_{dust} is the estimated dust luminosity of the (d12014, Draine et al. 2013) module from an energy balance, integrated over the full SED. L_{stellar} is estimated from the bc03 module. We give a summary of the mean values of these parameters for each sub-sample of the ADGD in Table 2.

3.2. Dimensionality Reduction of Host Galaxy Parameter Space

To search for connections between AGN selection method and host galaxy properties, we adopted the t-distributed stochastic neighbor embedding (t-SNE; van der Maaten & Hinton 2008) dimensionality reduc-

tion algorithm to uncover potential patterns and groupings in the data. This algorithm is based on the stochastic neighbor embedding technique (Hinton & Roweis 2002) which uses the sum of Kullback-Leibler divergences as a cost function while using a Student-t distribution instead of a Gaussian to compute the similarity of pair of point in the low-dimensional space. In addition to the construction of the cost function, the user has the ability to adjust the perplexity of this mapping, a hyper-parameter that defines a smooth measure of the number of neighbors for each point within the low-dimensional space. This algorithm is fairly sensitive to changes in perplexity. This method retains the local structure and reveals the global structure of the data within its low-dimensional map.

We input the redshift, host galaxy stellar mass, and Bayesian estimates of SFR, stellar metallicity, dust luminosity L_{dust} , AGN luminosity L_{AGN} , stellar luminosity L_{stellar} and excess E(B-V) color from the SED models. Figure 2 shows the results of this algorithm. We re-bin objects based on their position in the lower 2d dimensional map shown. We note that this method is fundamentally a data visualization tool, so to find significant grouping we iterate over a range of perplexities. We set the perplexity equal to 20 and the maximum iteration number to 1000. We also note that the definition of the groups found from this method are subjective; as we did this iterative process we found the global structure start to settle into groups that we defined visually. Moreover, the standard usage of t-SNE does not handle data uncertainties well. The use of uncertainties in the algorithm hinders its ability to preserve the local structure of the data. We opt to not include uncertainties within our t-SNE analysis. In summation, the classification of the populations within this low-dimensional map are subjective. In Section 4.2, we investigate the distribution of these properties in connection to the coverage of AGN selections for each grouping.

4. RESULTS

We present the results of the SED analysis here. We have modeled the SEDs of dwarf galaxies with a variety of AGN signatures found in W24, with 461 SEDs models making our quality cut on χ_r^2 . From these models, we collected Bayesian estimates for their host galaxy parameters. We draw connections between the subsamples and their host galaxy properties.

We used the t-SNE method to connect the array of selection techniques to the host galaxy parameters. t-SNE found nine new bins based on the similarity of the objects' host galaxy parameters. We then look at the properties for these new groups to investigate the in-

Median Sub-Sample Host Parameter values

Selection Method	Redshift	Mass [$M_{\odot}$]	SFR [$M_{\odot} / \text{yr}$]	$\log(L_{\text{AGN}})$ [erg/s]	$\log(L_{\text{dust}})$ [erg/s]	$\log(L_{\text{stellar}})$ [erg/s]	Stellar Metallicities	$E(B - V)$	f_{AGN}
BPT (AGN+Comps) Total: 197	$0.032^{+0.02}_{-0.011}$	$9.33^{+0.12}_{-0.29}$	$0.003^{+0.138}_{-0.003}$	$40.98^{+1.72}_{-2.15}$	$42.68^{+0.51}_{-0.33}$	$43.18^{+0.42}_{-0.35}$	$0.0008^{+0.0059}_{-0.0005}$	$0.0877^{+0.0146}_{-0.0078}$	$0.0019^{+0.1782}_{-0.0}$
[O I] Seyferts Total: 181	$0.034^{+0.039}_{-0.012}$	$9.21^{+0.24}_{-0.52}$	$0.008^{+0.352}_{-0.007}$	$40.19^{+2.3}_{-1.83}$	$42.66^{+0.62}_{-0.42}$	$43.13^{+0.45}_{-0.36}$	$0.0004^{+0.0048}_{-0.0002}$	$0.0875^{+0.0034}_{-0.0014}$	$0.0002^{+0.1034}_{-0.0}$
[O I] LINERS Total: 23	$0.028^{+0.016}_{-0.008}$	$9.31^{+0.15}_{-0.32}$	$0.003^{+0.177}_{-0.003}$	$39.28^{+2.53}_{-1.97}$	$42.48^{+0.54}_{-0.31}$	$42.92^{+0.38}_{-0.32}$	$0.0004^{+0.0065}_{-0.0002}$	$0.0875^{+0.0015}_{-0.0}$	$0.0001^{+0.0573}_{-0.0}$
[S II] Seyferts Total: 190	$0.036^{+0.03}_{-0.014}$	$9.24^{+0.21}_{-0.50}$	$0.005^{+0.13}_{-0.004}$	$40.13^{+2.25}_{-1.96}$	$42.67^{+0.47}_{-0.34}$	$43.13^{+0.39}_{-0.33}$	$0.0004^{+0.004}_{-0.0002}$	$0.0875^{+0.0025}_{-0.0013}$	$0.0002^{+0.0917}_{-0.0}$
[S II] LINERS Total: 39	$0.028^{+0.009}_{-0.006}$	$9.24^{+0.17}_{-0.42}$	$0.003^{+0.02}_{-0.002}$	$39.70^{+1.8}_{-2.89}$	$42.44^{+0.23}_{-0.33}$	$42.84^{+0.33}_{-0.39}$	$0.0004^{+0.006}_{-0.0002}$	$0.0875^{+0.0004}_{-0.0023}$	$0.0000^{+0.0341}_{-0.0}$
[He II] (AGN+Comps) Total: 241	$0.031^{+0.021}_{-0.001}$	$9.29^{+0.16}_{-0.44}$	$0.003^{+0.08}_{-0.003}$	$40.65^{+1.79}_{-2.24}$	$42.61^{+0.45}_{-0.37}$	$43.1^{+0.43}_{-0.35}$	$0.0005^{+0.0059}_{-0.0002}$	$0.0875^{+0.0029}_{-0.0115}$	$0.0004^{+0.1036}_{-0.0}$
Broad-line Total: 21	$0.047^{+0.05}_{-0.018}$	$9.26^{+0.18}_{-0.23}$	$0.109^{+1.458}_{-0.108}$	$43.08^{+0.62}_{-1.5}$	$43.18^{+0.58}_{-0.61}$	$43.53^{+0.49}_{-0.44}$	$0.0012^{+0.0071}_{-0.0008}$	$0.0877^{+0.0617}_{-0.073}$	$0.1919^{+0.6894}_{-0.0}$
WISE IR Total: 133	$0.055^{+0.025}_{-0.029}$	$8.83^{+0.44}_{-0.52}$	$0.144^{+1.143}_{-0.138}$	$41.10^{+1.97}_{-2.53}$	$43.07^{+0.65}_{-0.69}$	$43.41^{+0.54}_{-0.71}$	$0.0003^{+0.003}_{-0.0002}$	$0.0876^{+0.0229}_{-0.0011}$	$0.0002^{+0.2335}_{-0.0}$
X-ray (AGN) Total: 28	$0.031^{+0.016}_{-0.016}$	$9.28^{+0.14}_{-0.25}$	$0.007^{+0.375}_{-0.007}$	$41.51^{+1.47}_{-3.08}$	$42.72^{+0.63}_{-0.27}$	$43.15^{+0.52}_{-0.32}$	$0.0006^{+0.011}_{-0.0004}$	$0.0875^{+0.029}_{-0.0462}$	$0.000^{+0.2054}_{-0.0}$
Variability Total: 84	$0.034^{+0.02}_{-0.01}$	$9.24^{+0.18}_{-0.68}$	$0.006^{+0.112}_{-0.005}$	$40.43^{+2.01}_{-2.37}$	$42.68^{+0.52}_{-0.44}$	$43.1^{+0.47}_{-0.32}$	$0.0004^{+0.0066}_{-0.0002}$	$0.0875^{+0.0015}_{-0.0146}$	$0.000^{+0.0654}_{-0.0}$

Table 2. The median values of the host galaxy parameters estimated from the SED modeling. The total number of objects is the remaining size of the sub-sample once the χ^2_r cut was made. Upper and lower bounds given the width of 1 σ within each distribution.

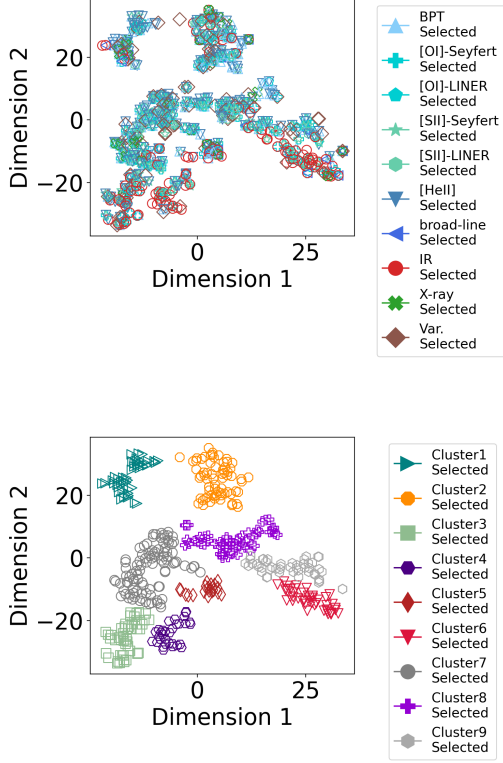


Figure 2. t-SNE clustering with a perplexity of 20, colored by AGN selection method (*top*) and after re-binning them into Clusters (*bottom*).

fluence of host galaxy further while trying to decouple their effects on galaxy selected by multiple methods.

The ADGD, including our sample, AGN selection results, and derived host galaxy and AGN parameters is available for public download on GitHub¹.

4.1. AGN Selection subsamples

Here we investigate host galaxy properties for AGN selected via each technique in the ADGD.

We first investigated the host galaxy stellar mass distribution of our subsamples. Figure 3 gives the cumulative sum of the selected fraction as a function of host galaxy stellar mass for each sub-sample. While most of the methods have comparable mass distributions, there are some interesting differences.

IR color selected a significantly higher fraction of low-mass candidates compared to other sub-populations, with variability having the second most bottom-heavy mass function. Comparatively, the BPT and broad-line

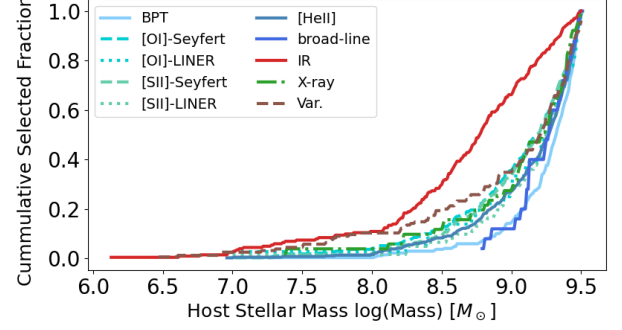


Figure 3. The cumulative fraction of objects selected as a function of host galaxy's stellar mass for each sub-sample identified in W24.

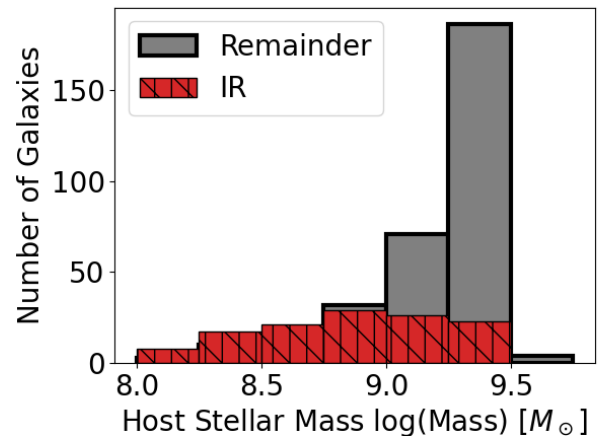
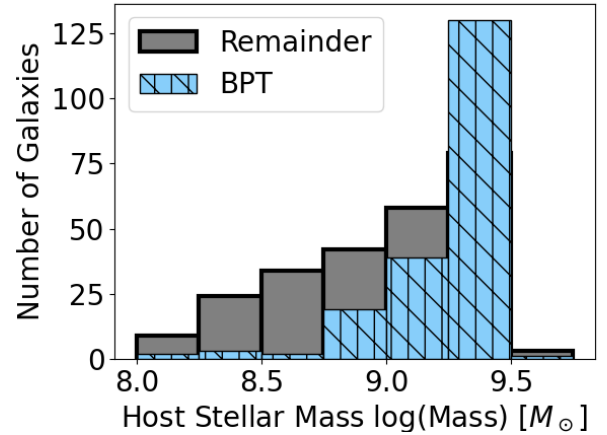


Figure 4. *Top:* Mass distribution of the BPT selected sub-sample in comparison to the Mass distribution of the remaining objects in the database not selected by that method (grey). *Bottom:* Same plot now separating out the IR selected sub-sample from the remaining database.

¹ <https://github.com/erikwasleske/Active-Dwarf-Galaxy-Database>

selections identify more massive candidates, suggesting these techniques might be biased towards AGN in more massive dwarf galaxies. These BPT and IR populations are compared to the remaining galaxies with the database in Figure 4. These plots demonstrate the mass biases. Applying a two-sample Kolmogorov-Smirnov (K-S) test (Massey 1951) to compare these sub-sample distributions to the remainder resulted in a near 0 p-values (much less than the stand threshold of 0.05), pointing to the stark difference in these distributions. We cautiously use this statistical test in comparing sub-sample host galaxy parameters as it relies on continuous data and is sensitive to outliers and extrema. We use it only as a reference guide to draw qualitative results.

The mass functions of the Seyfert and LINER populations, selected by either [O I] or [S II] emission, are in good agreement with each other.

We also look at the distributions of host galaxy and AGN parameters in connection with the selection technique to investigate the potential biases of each method individually. Table 2 lists the median Bayesian estimated parameters from our SED models. We additionally collect Bayesian estimates for the uncertainties of these values, which is the standard deviation of the distribution from all iterations of the model weighted by their probability (Boquien et al. 2019; Yang et al. 2022). Figure 5 presents histograms of the L_{AGN} , L_{dust} , L_{stellar} , and SFR estimates for galaxies selected via each technique. We quote mean Bayesian uncertainty for derived host galaxy parameters in Figures 5 and 7. The uncertainties are relatively small allowing us to draw conclusions about the differences in galaxy/AGN parameters between samples.

There are some differences in the L_{AGN} distributions of the subsamples. AGN identified via broad-line emission comprise the smallest sample (though these have the highest overlap with the other subsamples; W24). We find the broad-line AGN to have the highest median L_{AGN} by two orders of magnitude (median $L_{\text{AGN}} \approx 10^{43} \text{ erg s}^{-1}$) and the smallest spread of L_{AGN} . Furthermore, the broad-line selected population as the highest median L_{dust} and L_{stellar} . We show how the L_{AGN} distribution compares to the rest of the database in the top panel of Figure 6. The comparison of the [O I] and [S II] Seyfert distributions are given in the bottom two panels of Figure 6. These two-sample K-S test for these distributions are reporting near 0 p-value, with their shape pointing to a bias of less powerful AGN within these dwarfs compared to populations identified by the BPT, [He II], and IR diagnostics. We also see that [O I] and [S II] LINERS have the least luminous AGN (median

$L_{\text{AGN}} \approx 10^{39} \text{ erg s}^{-1}$). The remaining subsamples have median $L_{\text{AGN}} \approx 10^{40-41} \text{ erg s}^{-1}$.

There is overall little variation in the dust and stellar luminosities between subsamples. In particular, populations selected by narrow optical emission lines techniques have nearly identical dust luminosity distributions. IR and broad-line selected samples had slightly higher median dust luminosities. Stellar luminosities are all relatively similar (median $L_{\text{stellar}} \approx 10^{43} \text{ erg s}^{-1}$), though broad-line selected objects have slightly higher median stellar luminosities and LINERS have slightly lower.

We do note some key differences in SFRs between the subsamples, though in general each technique has a wide spread in SFR. The BPT, He II, and LINER subsamples have the lowest median SFRs (median SFR $\approx 0.003 \text{ M}_{\odot}/\text{yr}$). These median SFR values are approximately half that of the Variability and X-ray subsamples. Variability and X-ray selected AGN have higher medians; 0.006 and $0.007 \text{ M}_{\odot}/\text{yr}$, respectively. Broad-line AGN have median SFRs of $\sim 0.1 \text{ M}_{\odot}/\text{yr}$, and IR selected AGN have the highest median, at $0.14 \text{ M}_{\odot}/\text{yr}$.

In Figure 7, we present the histograms for the fractional contribution of the AGN IR luminosity to the total IR luminosity. This reflects the prevalence of the torus emission relative to the overall IR emission. The LINER, broad-line, X-ray and variability selected subsamples show a bimodal distribution. The broad-line subsample has the most galaxies with high AGN fraction ($f_{\text{AGN}} \geq 0.7$, ie. Carroll et al. 2021; Wasleske et al. 2022) with 11 of the 21 galaxies that meet this benchmark.

Next, we investigate the optical colors of our AGN subsamples. With the compiled photometry, we investigate the power-law slope of different aspects of the SEDs to discriminate between subsamples. Figure 8 shows the comparison of colors u-g and g-r using the de-reddened `cModelMag` from SDSS DR17. Immediately apparent is the blueness of IR selected AGN compared to their optical spectroscopic selected counterparts. In the u-g distributions, we see a separation between IR and spectroscopically selected population. The variability and X-ray selected AGN are more consistent with the other AGN populations. The IR selected population also has the biggest blue tail of its distribution for the g-r color. We can also see that the variability and X-ray g-r colors distributions are in between the IR and spectroscopically selected samples. The reddest objects are those selected with via the BPT and [He II] diagrams.

In Figure 9, we present the SFR - stellar mass relation for the AGN subsamples. We lay the AGN subsamples over a 2D histogram of $\sim 640,000$ galaxies from

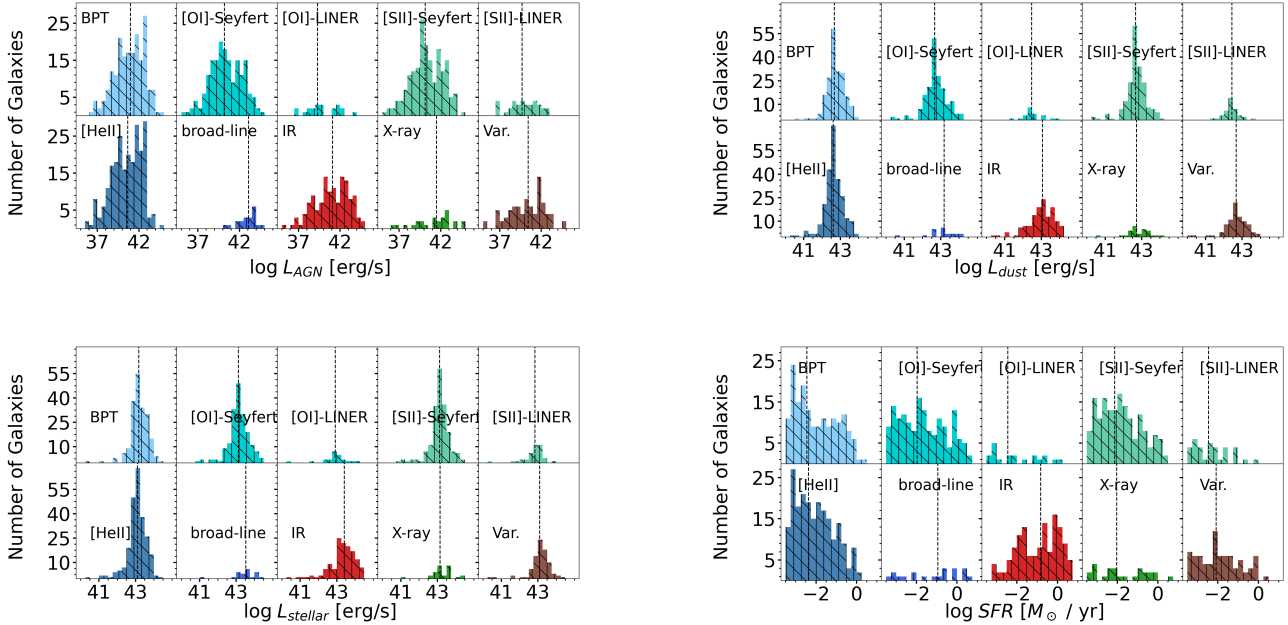


Figure 5. Histograms of L_{AGN} , L_{dust} , L_{stellar} , and SFR estimates for subsamples of each AGN selection method. Median values of each distribution are given shown by the dashed line. The SFR histograms (*bottom right*) is the logarithmic Bayesian estimated from **XCIGALE** for the SFR averaged over 100Myrs. The mean uncertainties for $\log L_{\text{AGN}}$, $\log L_{\text{dust}}$, $\log L_{\text{stellar}}$, and $\log \text{SFR}$ are 0.85, 0.05, 0.04, 0.31 respectively.

GALEX-SDSS-WISE Legacy Catalog Shallow ‘All-sky’ catalog (GSWLC-2, [Salim et al. 2016, 2018](#))². The SFR and stellar masses of these galaxies are Bayesian estimates from SED modeling. These galaxies, with coverage across the UV-optical-IR, give a large population against which we can compare the physical properties of AGN in low-mass galaxies. Within the SFR-stellar mass space, we can clearly identify the blue cloud of star-forming galaxies and red sequence of quiescent galaxies ([Daddi et al. 2007](#); [Rodighiero et al. 2011](#)). We note the wide distribution of the SFR for the active dwarf galaxies, with the IR selected population matching best to the low-mass regime of the blue cloud of star-forming galaxies. Those that are optically selected tend towards the red quiescent sequence of galaxies with lower SFR.

A large fraction of the distribution of the optical, variability, and IR selected lay within the low-mass regime of the green valley, transitioning from eras of high star-formation to becoming ‘red and dead’, whereas the X-ray selected population tends towards the bimodal distributions of the blue-cloud and red sequence.

4.2. *t*-SNE Clustering

Below we discuss the results from the machine learning technique described in Section 3 to connect the host

galaxy values to the diversity of AGN signatures found in W24.

From the implementation of *t*-SNE in Figure 2, we separated the 2D map into the nine clusters shown. We analyze the distribution of the input parameters listed in Section 3.2. Table 3 gives the number of galaxies within each group and the median values for these parameters for each cluster. In Figure 10 we show the distributions of these parameters within Cluster 6. The bottom right panel of the Figure shows the fractional value of each AGN selection technique within this cluster. Cluster 6 is dominated by those selected via IR colors with a fair overlap with Seyferts selected from the [O I] and [S II] diagrams. In the Appendix, Figures 14, 15, and 16 show the distributions for the remaining clusters.

In lieu of statistical values to determine the spread of the distribution of each parameter, we instead compare the shape of the distributions by evaluating the difference of the parameter value from the median value of its distribution for percentile steps within said distribution. We opt for this evaluation as distributions of the host galaxy parameters for each cluster deviate from normal distributions, with some having gaps within distribution. These plots thus give us a better understanding of the overall shape and dispersion of the values within. An example plot for L_{AGN} is given in Figure 11. It shows the relation between the percentile steps through the

² <https://salims.pages.iu.edu/gswlc/>

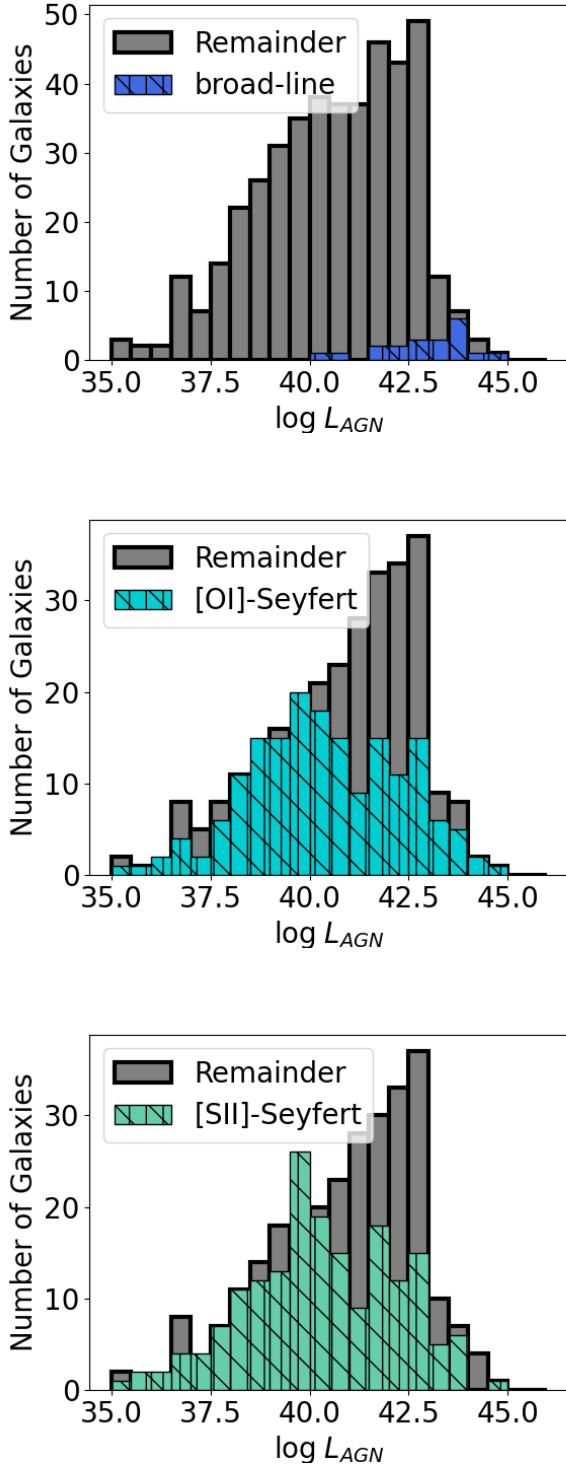


Figure 6. L_{AGN} distribution of the broad-line selected sub-sample (*top*), [O I] (*middle*) and [S II] (*bottom*) Seyferts compared to the remaining galaxies in the database.

distribution of $\log(L_{AGN})$ for each cluster, compared to the difference of value in $\log(L_{AGN})$ at that percentile to

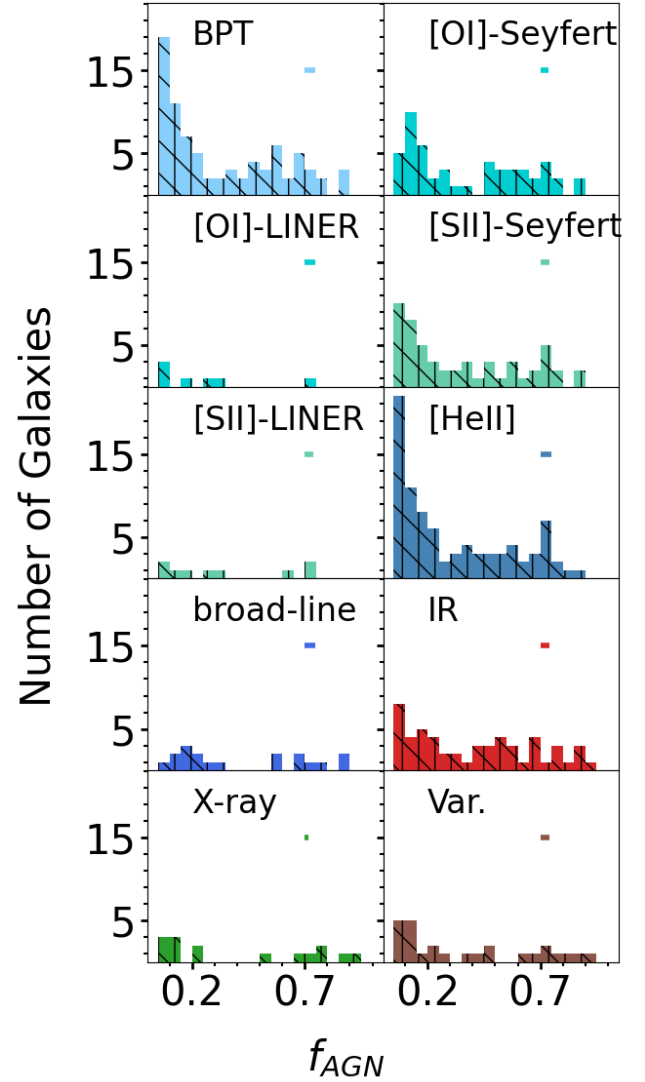


Figure 7. Histogram of fractional contribution of the AGN template to the IR luminosity, f_{AGN} , for each AGN selection method subsample. The color matched horizontal lines give the mean Bayesian estimate for the uncertainty for the values in each histogram, with mean error being 0.03.

the median value of the distribution. Analogous figures for the remaining galaxy parameters are shown in the Appendix in Figures 17, 18, 19, and 20. Within these plots, heavier weights in distant bins from the median result in steeper slopes. Narrow distributions are represented by lines with slopes approaching zero.

The t-SNE method was successful in finding groups dominated by optical narrow emission line selections (ie. Clusters 1, 2, 7, and 8) and those dominated by IR color selection (Clusters 4 and 6). Distinguishing significant contribution of X-ray and variability selected

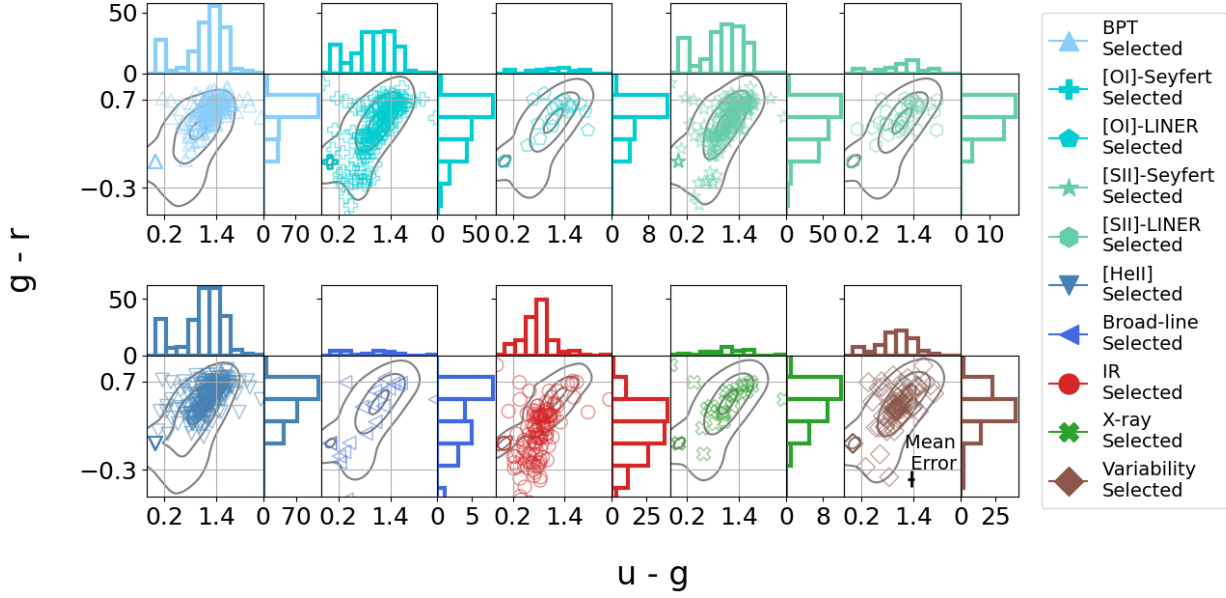


Figure 8. SDSS $u-g$ vs $g-r$ diagram for the subsamples of AGN selected population shown on each panel. The grey contours show the 10, 50, and 90 percentile of the database *not* selected by corresponded method for that panel (i.e. the Remainder of the database). The mean error of these color measurements for the full sample is shown in black in the last panel.

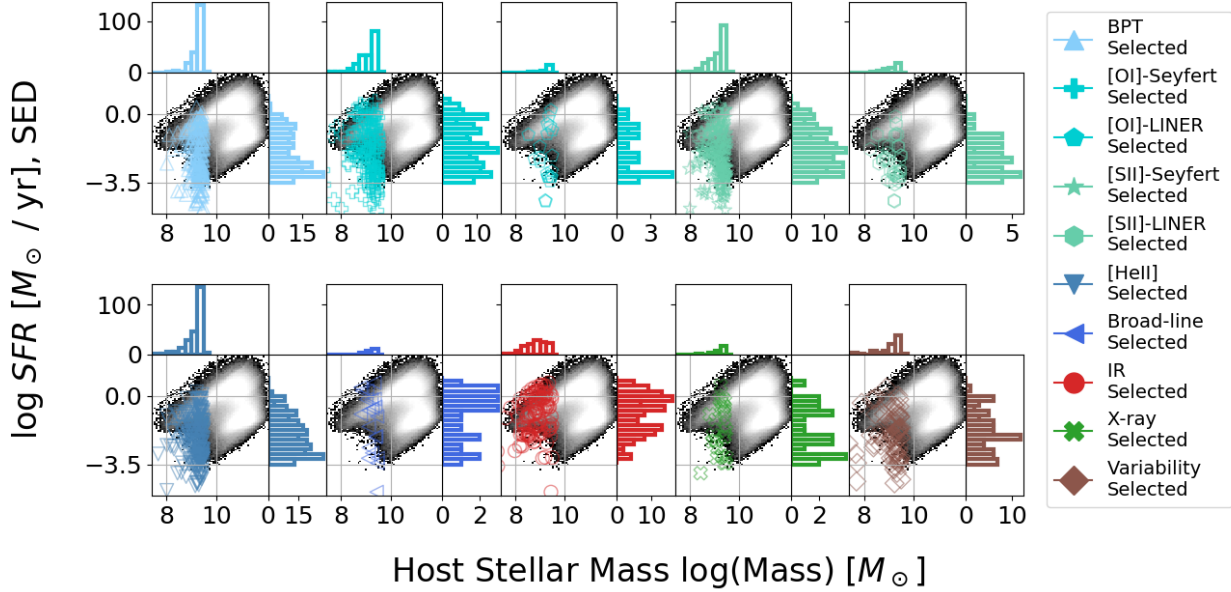


Figure 9. We show the the SFR-Mass relation for the sub-populations of the 461 AGN in our analysis within each panel. In the background is a 2-D histogram of $\sim 640,000$ galaxies within the GALEX-SDSS-WISE Legacy Catalog. We compare using our Bayesian estimates of average SFR over 100Myrs. The median uncertainty is $\sigma_{\log \text{SFR}} = 0.09 [M_{\odot}/\text{yr}]$.

AGN within each cluster is limited by the smaller population of objects and thus AGN selection of galaxies within the ADGD. Those clusters with the highest variability selected fractions also have [He II] selected fractions of $\sim 50\%$. We note the fraction of X-ray selected AGN scales with the fraction of BPT selected within each cluster.

5. DISCUSSION

We have estimated host galaxy and AGN properties for a set of active dwarf galaxies that were selected by a variety of electromagnetic accretion signatures. In evaluating the distribution and relations of these parameters for each sub-sample of dwarf AGN, we attempt to draw conclusions about the biases of each selection method. Then, we investigated the parameters distributions within the clusters of the t-SNE results to search for trends within the overlap of the selection methods. In the following sections, we discuss the implications of these results.

5.1. Different evolutionary states of AGN subsamples

Overall, the results we find from the distribution of host galaxy parameters for each sub-sample population provide quantitative support for previous qualitative understanding of the biases of each selection technique (Mendez et al. 2013; Hainline et al. 2016; Cann et al. 2019; Burke et al. 2021; Arcodia et al. 2024). We find that our IR sub-sample contains high levels of star formation and is generally bluer in the optical. The optical spectroscopically selected AGN reside in the host galaxies with the lowest average SFRs in the sample. Our broad-line and X-ray samples contain relatively luminous AGN compared to the other subsamples. We can now discuss these results within the perspective of other studies.

Previous works have examined the selection of AGN in connection with the mass and age of the host galaxy, and the accretion process (Silverman et al. 2008; Schawinski et al. 2009; Hickox et al. 2009). Hickox et al. (2009) evaluated three different classes of non-dwarf, massive AGN, identified by X-ray, IR, and Radio emissions, at a redshift of $0.25 < z < 0.8$. They found that the X-ray selected objects were in galaxies of all colors whereas the IR selected group comprised bluer and less luminous galaxies. The radio population was found to have very low Eddington ratios and the most massive BHs. The distribution parameters of these three classes points to an evolutionary scenario. They suggest that IR AGN are at an earlier phase of AGN-galaxy co-evolution, accreting at higher rates with smaller central BHs and higher SFR host galaxies than those selected by X-ray, with

radio being the latest stage. Radio galaxies were found to have the most massive halos, with IR selected having the least massive.

Ellison et al. (2016) investigated host galaxy SFRs of AGN selected optically, from mid-IR, and from radio emission, finding a high dependence of the SFR on how the AGN is selected, especially for obscured systems. This supports an evolutionary sequence of post-merger systems having a dependence on obscuration to increased SFRs. Their results show that merger events serve as a catalyst for extreme SF and high accretion rates for absorbed AGN. They find enhanced SFR from the mid-IR selected AGN population and obscured galaxies in the X-ray. They conclude that IR selected AGN have the highest fraction of mergers within their study as identified in Galaxy Zoo. For an unobscured population, the results are less clear, as the un-obscured optically-selected AGN lack enhanced SFRs and are unlikely to be fueled by recent merger events.

With our photometric data and estimated values of the host galaxy parameters, we can discuss our dwarf AGN subsamples within this evolutionary scheme. The star formation properties, stellar luminosity, and mass distribution suggests that the IR sub-sample is at an earlier evolutionary phase in comparison to our other populations. Figure 12 shows the magnitude color space of the subsamples. For similar distributions of r-band magnitude, there is evidence an evolution of IR-to-variability-to-optically selected with increasing u-r color, confirming the results of the SFR-Mass plot of Figure 9. We find greater separation between our X-ray and IR populations than the X-ray and IR selected populations of massive AGN with extended optical counterparts plotted on a similar diagram of Section 5 Hickox et al. (2009). Hickox et al. (2009) quotes their X-ray population as having mean magnitude and dispersion ($\langle M_r \rangle, \sigma_{M_r}$) values of $(-21.3, 0.7)$ with mean color and dispersion ($\langle u-r \rangle, \sigma_{u-r}$) of $(2.2, 0.4)$, and their IR population having values ($\langle M_r \rangle, \sigma_{M_r}$) = $(-21.1, 0.6)$ and ($\langle u-r \rangle, \sigma_{u-r}$) = $(2.1, 0.4)$. Comparatively, our X-ray population has values of ($\langle M_r \rangle, \sigma_{M_r}$) = $(-19.8, 4.6)$ and ($\langle u-r \rangle, \sigma_{u-r}$) = $(1.5, 0.7)$. Our IR populations values are ($\langle M_r \rangle, \sigma_{M_r}$) = $(-18.9, 3.9)$ and ($\langle u-r \rangle, \sigma_{u-r}$) = $(0.8, 0.8)$. The distance between our X-ray and IR mean values within this color-magnitude space is 1.14, much greater than the separation of 0.22 between the median values of the massive populations from Hickox et al. (2009). Overall our populations, being at lower redshift and mass, are less luminous and bluer. While our populations are more dispersed over this space, the trend of the massive X-ray populations being redder holds true and is more apparent than in the massive galaxy samples. This evi-

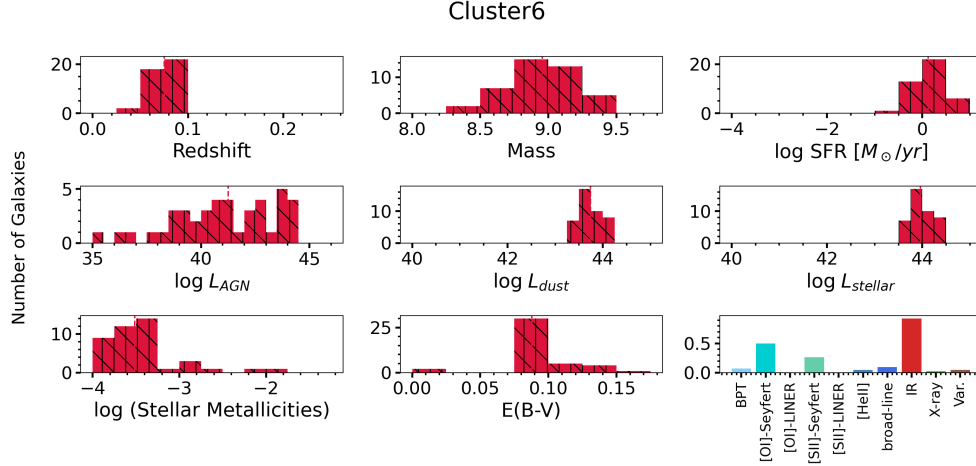


Figure 10. Distribution of host galaxy parameters for Cluster 6 from the t-SNE dimensionality reduction.

Median Cluster Host Parameter Values

Cluster Number	Redshift	Mass [$M_{\odot}$]	SFR [$M_{\odot} / \text{yr}$]	$\log(L_{\text{AGN}})$ [erg/s]	$\log(L_{\text{dust}})$ [erg/s]	$\log(L_{\text{stellar}})$ [erg/s]	Metallicities	$E(B - V)$
Cluster 1 Total: 46	0.0303	9.30	0.0074	42.44	42.57	43.08	0.0016	0.0092
Cluster 2 Total: 68	0.0332	9.37	0.0089	41.43	42.57	43.09	0.0075	0.0879
Cluster 3 Total: 49	0.0246	8.4	0.0045	38.82	42.34	42.70	0.0004	0.0875
Cluster 4 Total: 32	0.0568	8.47	0.0841	40.78	42.90	43.25	0.0002	0.08772
Cluster 5 Total: 20	0.0265	9.20	0.03820	42.86	42.91	43.38	0.0027	0.1491
Cluster 6 Total: 42	0.0752	8.96	1.3177	41.25	43.74	43.96	0.0003	0.0878
Cluster 7 Total: 97	0.0288	9.2	0.0014	39.20	42.55	43.015	0.0004	0.0875
Cluster 8 Total: 59	0.0439	9.4	0.0090	40.41	42.96	43.42	0.0003	0.0875
Cluster 9 Total: 48	0.1059	9.28	0.1847	42.03	43.26	43.58	0.0009	0.0877

Table 3. The median values of the host galaxy parameters estimated from the SED modeling for the clusters defined by the t-SNE dimensionality reduction. The spread of these distributions are given by the percentile plots discussed in Section 4.2.

dence supports the evolutionary model of AGN in that our X-ray subsample is an older redder phase, existing later in the green valley than the IR subsample.

5.2. Are IR selected Dwarf AGN just Star Forming Galaxies?

To investigate the concern of highly star-forming galaxies masquerading as AGN in IR color selections (Satyapal et al. 2018), we seek to understand this sub-

sample better. As stated in W24, 457/702 of the AGN within the ADGD was selected by more than one method. Of the 201 dwarf galaxies selected as AGN from IR colors, 130 objects were selected by at least one additional method. Figure 13 breaks the SFR histogram of the IR selected sub-sample into those selected by another method from those only selected from IR colors. Of the 71 objects selected by IR colors only, 47 have non-zero fractional contribution of the AGN template

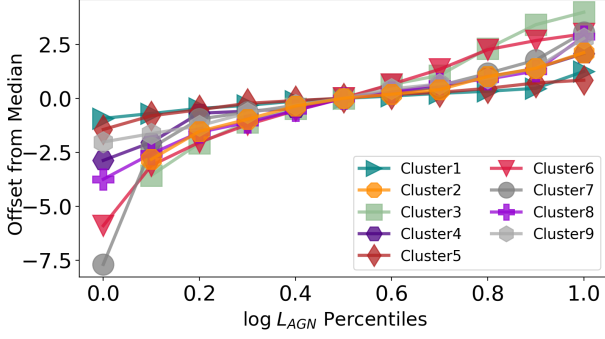


Figure 11. Percentile of the distribution vs. offset from median for each cluster within the t-SNE analysis for L_{AGN} .

to their SED. However, the median AGN fraction of these 47 objects is 0.02, with the maximum value being 0.66 in this subset. The remaining 24 objects do not have SED modeling that met our criteria χ_r^2 mentioned in Section 3.1. As mentioned above, these objects were not included in our analysis. The small non-zero AGN SED contribution for the remaining 47 IR-only selected objects adds to ambiguity surrounding the purity of the IR selection. As the remaining IR-only group contains very few objects with low SFRs, this points to this method’s bias of identifying high star-forming dwarf galaxies. Additionally, the contribution of dust within the IR selected sub-sample in unison with the stellar emission and SFR shows the caution needed when implementing IR selection as young star-forming dwarfs can mimic AGN colors (Hainline et al. 2016).

5.3. AGN Signatures and Host Parameter Distributions of the t-SNE Groups

The re-binning of groups based on AGN selection techniques (W24) to similarities in host galaxy parameters via t-SNE opens a new avenue to examine the disagreement between AGN selection methods. This dimensionality reduction method allowed us to explore the intersections of the Venn diagram of the AGN selection techniques applied to the ADGD.

The two most distinct groupings in this low-dimensional map are Clusters 1 and 2. These two assemblies have the most spread distributions of L_{dust} and L_{stellar} . Both groups contain galaxies selected via a variety of methods, with the most prominent technique being [He II] selection. Cluster 1 also has distinctively low $E(B-V)$ compared to the high $E(B-V)$ of Cluster 2. This points to [He II] being less sensitive to the influences of dust and stellar emission in active dwarfs.

We find that for Clusters 1, 2, 7, and 8 which have the highest fraction of spectroscopically selected objects all have similar medians and distributions of L_{stellar} . Clus-

ters 2, 7 and 8 have lower median L_{AGN} and wider distributions of L_{AGN} than Cluster 1 (Figure 11). Of these groups, Cluster 1 is the most purely [He II] selected while having the third highest stellar metallicities. This is consistent with the [He II] being more sensitive to the hardness of the ionizing source (Sartori et al. 2015).

Clusters 4 and 6 have a significant fraction of IR selected ($> 65\%$) objects. The second and third most significant selection method within both of these groups are Seyferts from the [O I] and [S II] diagrams. The most apparent differences between these two clusters is the wide distribution of SFR and L_{AGN} for Cluster 6. Comparing to the other seven groups, Cluster 4’s separation is dominated by the low mass of its objects.

Clusters 3, 5, and 9 are the groupings with the most diverse contributions of AGN selection methods. Cluster 3 lacks the contribution from BPT selection that Clusters 5 and 9 have, while having the highest contribution from variability selected of the three groups. Cluster 5 has significant selection fraction as BPT, Seyfert, and IR. This population has the highest obscuration among the clusters.

In summary, the binning from t-SNE can be used to investigate the overlap of AGN selection techniques. For the spectroscopically dominated clusters (1,2,7,8) compared the remaining groups, there is an inherent bias of spectroscopic selection towards selecting low SF, higher mass candidates than those selected via photometric methods. No cluster corresponds to a single selection method. The assortment of sub-sets within each group can provide insight into the scatter within the parameter spaces utilized by each selection technique. The lack of clean division of selection subsamples within the t-SNE clusters reflects the inherent degeneracies in computing host galaxy parameters and the difficulty of disentangling different selection effects.

For example, in looking at Figure 7 of W24, we see that variability selected subsample is scattered across the [He II] diagram, with a fraction of the subset lacking strong [He II] and [N II] emission. Within our t-SNE analysis, Cluster 3 has strong contributions of the variability and [He II] samples within its populations, pointing to a possible underlying cause for this separation in the variable population on the [He II] diagram. Cluster 3 represents a population that have less luminous AGN, and are forming less stars compared to the variability sub-sample as a whole. However, the diverse contributions of the selection subsamples within each Cluster contaminate the ability to infer a connection, as Cluster 3 also has a high IR selected population, shifting the stellar and dust luminosity distributions higher and the mass distribution lower.

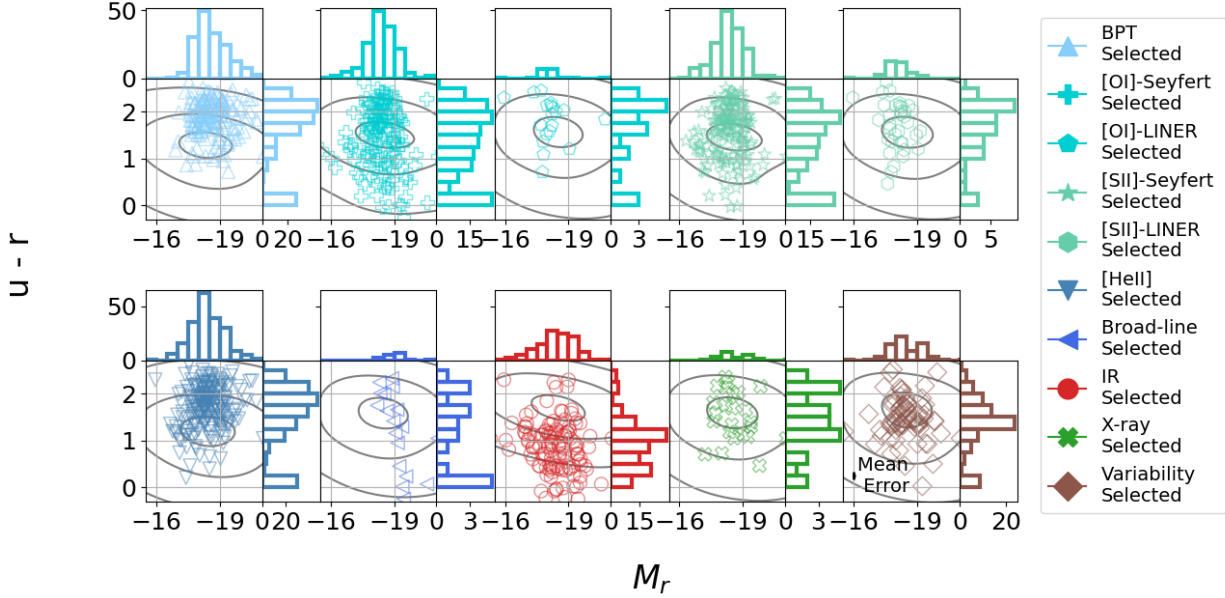


Figure 12. Magnitude-color space of absolute magnitude in r band for each subsample, calculated from the `cModelMag` vs. the `cModelMag` u-r color. Grey contours and mean error bar are the same as in Figure 8.

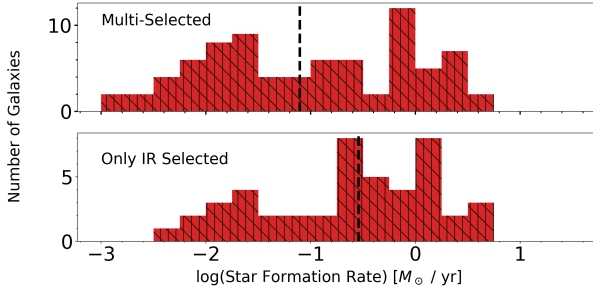


Figure 13. The histogram of SFR for the IR selected subsample of those selected by many method (*top*) and those identified by their IR emissions (*bottom*) alone. The vertical dashed lines give the median values of each distribution.

Comparatively, Cluster 8 contains heavy contributions from Variability and [He II] but lacks an IR population in favor of higher BPT contribution. From this, we see a brighter AGN luminosity, even lower SFR and a more massive sample than in Cluster 3. Looking at the [He II] diagram of Figure 8 in W24, the scatter maybe do to the AGN luminosity and SFR, as the BPT selected population occupies a region of higher [He II] and [N II] emission compared to those Variability objects that are selected on this diagram. To get a full understanding of this interplay, a full analysis of the location of each object within each parameter space matched to the t-SNE cluster would be required. The degeneracy and contamination of all selected sub-samples contributing

to all t-SNE clusters leads to the need to view of this problem on an object-to-object basis.

5.4. Need for New SED Templates for Dwarf AGN

Over the course of this work, we analyzed the results of host galaxy parameters estimated from SED models. However, it is important to recall that these templates were constructed for more massive host galaxies and their active nuclei. The results presented are dependent on the accuracy of the decomposition of these SED models into the intrinsic emission components in the galaxy. The difference between an accurate dwarf AGN template and the massive AGN templates used is unclear.

The construction of a composite dwarf AGN SED template requires high-resolution photometry of active dwarf galaxies. We have discussed the difference in the SED for our various subsamples in this study, thus a focus on unobscured, bright dwarf AGN would provide the most ideal population to build this initial dwarf AGN template.

6. CONCLUSIONS

Using the database of active dwarf galaxies compiled in W24, we investigated the connection between their accretion signatures and host galaxy properties. Of the 702 AGN within the ADGD, we develop 461 statistically significant SED models from which to estimate host galaxy parameters. We then attempted to re-bin

objects based on these properties. We gathered additional photometric data to construct SED models using XCIGALE (Yang et al. 2022). Our models provided Bayesian estimates of the host galaxy properties. These values were used to draw connections to the AGN selection subsamples as well as serving as input for t-SNE to take the high-dimensional parameter space of redshift, host galaxy stellar mass, SFR, stellar metallicity, L_{dust} , L_{AGN} , L_{stellar} and E(B-V) color to develop a low dimensional representation of the ADGD. Our conclusions are as follows:

- The cumulative mass function shows IR and Variability subsamples biased towards AGN in less massive galaxies, whereas the BPT and broad-line selections are biased towards more massive active dwarf galaxies.
- The broad-line subsample has the brightest median L_{AGN} , while LINER objects have the lowest median L_{AGN} . The [O I] and [S II] Seyfert populations have a distinctly different L_{AGN} distributions from the rest of the database.
- The L_{dust} distributions are similar in median values and variations across subsamples.
- The BPT, [He II], and LINER subsamples have a median SFR that is half of the median value for the Variability and X-ray subsamples.
- Our results within magnitude-color space suggest agreement with an evolutionary picture of our subsamples. This defines our IR sample as being in an earlier phase and the optically selected objects representing the later stages of this evolution. The X-ray and variability populations serve as the middle ground within this interpretation.
- Future studies into the morphologies of these systems can be used to investigate the merger histories of these populations to search for further evidence of this evolution.
- t-SNE provided a visualization of the ADGD within a low-dimensional parameter space to try to find connections between objects selected as AGN from multiple methods. However, the degeneracies of these populations does not lead to a fulfilling conclusions outside of re-evaluation of the host parameters in connection to a new binning schematic.
- In evaluating our results for the most complete suite of AGN selection methods, we note that if

only [He II] and IR selections are applied, 73% of our database would be recovered, the highest rate of recovery for any combination of two methods. Throughout our analysis, we noted the stark differences between populations selected by these methods, along with t-SNE clustering emphasizing the difference between the populations selected by spectroscopy and IR. Furthermore, the combination of [He II], IR and Variability finds 88.3% of the total database, the highest recovery rate of any suite of three selection methods used.

Overall, we find evidence for evolutionary differences between the subsamples of our dwarf galaxy database. The degeneracies of these populations starts to muddy these results. However, we note that this work is highly dependent upon the accuracy of our SED models and its decomposition. Currently, there exists no empirical model of a broadband SED for active dwarf galaxies.

Decomposition of high resolution imaging of active dwarf galaxies across the electromagnetic spectrum can help to disentangle these issues and provided a more accurate prescription for the SED models of dwarf galaxies.

The underlying data set used in this work and Wasleske & Baldassare (2024) is available in Zenodo at doi: 10.5281/zenodo.17407817. The catalog is in a FITS format and contains positions, redshifts, photometry, emission line luminosities, SED fitting parameters, and AGN selection indicators.

This work is supported by the NASA Astrophysics Database Analysis Program through Grant 80NSSC23K0460.

We thank the anonymous referee for comments that have improved this manuscript.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration.

This research has made use of data obtained from the Chandra Data Archive and the Chandra Source Catalog, and software provided by the Chandra X-ray Center (CXC) in the application packages CIAO and Sherpa.

This research has made use of data obtained from the 4XMM XMM-Newton serendipitous source catalog compiled by the 10 institutes of the XMM-Newton Survey Science Centre selected by ESA.

Software: Astropy (Astropy Collaboration et al. 2018), `astroquery` (Ginsburg et al. 2019), Matplotlib (Hunter 2007), NumPy (Harris et al. 2020), Scikit-learn (Pedregosa et al. 2011).

REFERENCES

- Abdurro'uf, Accetta, K., Aerts, C., et al. 2022, *ApJS*, 259, 35, doi: [10.3847/1538-4365/ac4414](https://doi.org/10.3847/1538-4365/ac4414)
- Antonucci, R. 1993, *ARA&A*, 31, 473, doi: [10.1146/annurev.aa.31.090193.002353](https://doi.org/10.1146/annurev.aa.31.090193.002353)
- Arcodia, R., Merloni, A., Comparat, J., et al. 2024, *A&A*, 681, A97, doi: [10.1051/0004-6361/202347531](https://doi.org/10.1051/0004-6361/202347531)
- Assef, R. J., Stern, D., Noirot, G., et al. 2018, *ApJS*, 234, 23, doi: [10.3847/1538-4365/aaa00a](https://doi.org/10.3847/1538-4365/aaa00a)
- Assef, R. J., Kochanek, C. S., Brodwin, M., et al. 2010, *ApJ*, 713, 970, doi: [10.1088/0004-637X/713/2/970](https://doi.org/10.1088/0004-637X/713/2/970)
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, 156, 123, doi: [10.3847/1538-3881/aabc4f](https://doi.org/10.3847/1538-3881/aabc4f)
- Baldassare, V. F., Geha, M., & Greene, J. 2018, *ApJ*, 868, 152, doi: [10.3847/1538-4357/aae6cf](https://doi.org/10.3847/1538-4357/aae6cf)
- Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, *PASP*, 93, 5, doi: [10.1086/130766](https://doi.org/10.1086/130766)
- Berk, D. E. V., Wilhite, B. C., Kron, R. G., et al. 2004, *The Astrophysical Journal*, 601, 692, doi: [10.1086/380563](https://doi.org/10.1086/380563)
- Bianchi, L., Shiao, B., & Thilker, D. 2017, *ApJS*, 230, 24, doi: [10.3847/1538-4365/aa7053](https://doi.org/10.3847/1538-4365/aa7053)
- Birchall, K. L., Watson, M. G., & Aird, J. 2020, *Monthly Notices of the Royal Astronomical Society*, 492, 2268, doi: [10.1093/mnras/staa040](https://doi.org/10.1093/mnras/staa040)
- Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, *A&A*, 622, A103, doi: [10.1051/0004-6361/201834156](https://doi.org/10.1051/0004-6361/201834156)
- Brandt, W., & Hasinger, G. 2005, *Annual Review of Astronomy and Astrophysics*, 43, 827, doi: [10.1146/annurev.astro.43.051804.102213](https://doi.org/10.1146/annurev.astro.43.051804.102213)
- Bruzual, G., & Charlot, S. 2003, *Monthly Notices of the Royal Astronomical Society*, 344, 1000, doi: [10.1046/j.1365-8711.2003.06897.x](https://doi.org/10.1046/j.1365-8711.2003.06897.x)
- Burke, C. J., Liu, X., Chen, Y.-C., Shen, Y., & Guo, H. 2021, *Monthly Notices of the Royal Astronomical Society*, 504, 543, doi: [10.1093/mnras/stab912](https://doi.org/10.1093/mnras/stab912)
- Burke, C. J., Shen, Y., Blaes, O., et al. 2021, *Science*, 373, 789, doi: [10.1126/science.abg9933](https://doi.org/10.1126/science.abg9933)
- Burke, C. J., Liu, X., Shen, Y., et al. 2022, *MNRAS*, 516, 2736, doi: [10.1093/mnras/stac2262](https://doi.org/10.1093/mnras/stac2262)
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *The Astrophysical Journal*, 533, 682, doi: [10.1086/308692](https://doi.org/10.1086/308692)
- Cann, J. M., Satyapal, S., Abel, N. P., et al. 2019, *The Astrophysical Journal*, 870, L2, doi: [10.3847/2041-8213/aaf88d](https://doi.org/10.3847/2041-8213/aaf88d)
- Carroll, C. M., Hickox, R. C., Masini, A., et al. 2021, *The Astrophysical Journal*, 908, 185, doi: [10.3847/1538-4357/abd185](https://doi.org/10.3847/1538-4357/abd185)
- Cattaneo, A., Faber, S. M., Binney, J., et al. 2009, *Nature*, 460, 213, doi: [10.1038/nature08135](https://doi.org/10.1038/nature08135)
- Chabrier, G. 2003, *Publications of the Astronomical Society of the Pacific*, 115, 763, doi: [10.1086/376392](https://doi.org/10.1086/376392)
- Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, *arXiv e-prints*, arXiv:1612.05560, doi: [10.48550/arXiv.1612.05560](https://doi.org/10.48550/arXiv.1612.05560)
- Daddi, E., Dickinson, M., Morrison, G., et al. 2007, *The Astrophysical Journal*, 670, 156, doi: [10.1086/521818](https://doi.org/10.1086/521818)
- Desroches, L.-B., Greene, J. E., & Ho, L. C. 2009, *ApJ*, 698, 1515, doi: [10.1088/0004-637X/698/2/1515](https://doi.org/10.1088/0004-637X/698/2/1515)
- Draine, B. T., & Li, A. 2007, *ApJ*, 657, 810, doi: [10.1086/511055](https://doi.org/10.1086/511055)
- Draine, B. T., Aniano, G., Krause, O., et al. 2013, *The Astrophysical Journal*, 780, 172, doi: [10.1088/0004-637X/780/2/172](https://doi.org/10.1088/0004-637X/780/2/172)
- Ellison, S. L., Teimoorinia, H., Rosario, D. J., & Mendel, J. T. 2016, *MNRAS*, 458, L34, doi: [10.1093/mnras/slw012](https://doi.org/10.1093/mnras/slw012)
- Ferrarese, L., & Merritt, D. 2000, *ApJL*, 539, L9, doi: [10.1086/312838](https://doi.org/10.1086/312838)
- Flewelling, H. A., Magnier, E. A., Chambers, K. C., et al. 2020, *ApJS*, 251, 7, doi: [10.3847/1538-4365/abb82d](https://doi.org/10.3847/1538-4365/abb82d)
- Gebhardt, K., Bender, R., Bower, G., et al. 2000, *ApJL*, 539, L13, doi: [10.1086/312840](https://doi.org/10.1086/312840)
- Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., et al. 2019, *AJ*, 157, 98, doi: [10.3847/1538-3881/aafc33](https://doi.org/10.3847/1538-3881/aafc33)
- Greene, J. E., & Ho, L. C. 2007, *ApJ*, 656, 84, doi: [10.1086/509064](https://doi.org/10.1086/509064)
- Greene, J. E., Strader, J., & Ho, L. C. 2020, *Annual Review of Astronomy and Astrophysics*, 58, 257, doi: [10.1146/annurev-astro-032620-021835](https://doi.org/10.1146/annurev-astro-032620-021835)
- Groves, B. A., Heckman, T. M., & Kauffmann, G. 2006, *Monthly Notices of the Royal Astronomical Society*, 371, 1559, doi: [10.1111/j.1365-2966.2006.10812.x](https://doi.org/10.1111/j.1365-2966.2006.10812.x)
- Hainline, K. N., Reines, A. E., Greene, J. E., & Stern, D. 2016, *The Astrophysical Journal*, 832, 119, doi: [10.3847/0004-637x/832/2/119](https://doi.org/10.3847/0004-637x/832/2/119)
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, 585, 357, doi: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2)
- Hickox, R. C., Jones, C., Forman, W. R., et al. 2009, *The Astrophysical Journal*, 696, 891, doi: [10.1088/0004-637X/696/1/891](https://doi.org/10.1088/0004-637X/696/1/891)
- Hinton, G. E., & Roweis, S. 2002, in *Advances in Neural Information Processing Systems*, ed. S. Becker, S. Thrun, & K. Obermayer, Vol. 15 (MIT Press). https://proceedings.neurips.cc/paper_files/paper/2002/file/6150ccc6069bea6b5716254057a194ef-Paper.pdf
- Ho, L. C., Filippenko, A. V., & Sargent, W. L. W. 1997, *The Astrophysical Journal Supplement Series*, 112, 315, doi: [10.1086/313041](https://doi.org/10.1086/313041)

- Hunter, J. D. 2007, *Computing in Science & Engineering*, 9, 90, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55)
- Hviding, R. E., Hainline, K. N., Rieke, M., et al. 2022, *The Astronomical Journal*, 163, 224, doi: [10.3847/1538-3881/ac5e33](https://doi.org/10.3847/1538-3881/ac5e33)
- Jarrett, T. H., Cohen, M., Masci, F., et al. 2011, *The Astrophysical Journal*, 735, 112, doi: [10.1088/0004-637x/735/2/112](https://doi.org/10.1088/0004-637x/735/2/112)
- Just, D. W., Brandt, W. N., Shemmer, O., et al. 2007, *ApJ*, 665, 1004, doi: [10.1086/519990](https://doi.org/10.1086/519990)
- Kewley, L. J., Groves, B., Kauffmann, G., & Heckman, T. 2006, *Monthly Notices of the Royal Astronomical Society*, 372, 961, doi: [10.1111/j.1365-2966.2006.10859.x](https://doi.org/10.1111/j.1365-2966.2006.10859.x)
- Kormendy, J., & Ho, L. C. 2013, *ARA&A*, 51, 511, doi: [10.1146/annurev-astro-082708-101811](https://doi.org/10.1146/annurev-astro-082708-101811)
- Kozłowski, S. 2017, *A&A*, 597, A128, doi: [10.1051/0004-6361/201629890](https://doi.org/10.1051/0004-6361/201629890)
- Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, *MNRAS*, 379, 1599, doi: [10.1111/j.1365-2966.2007.12040.x](https://doi.org/10.1111/j.1365-2966.2007.12040.x)
- Lehmer, B. D., Eufrasio, R. T., Basu-Zych, A., et al. 2021, *ApJ*, 907, 17, doi: [10.3847/1538-4357/abcec1](https://doi.org/10.3847/1538-4357/abcec1)
- Lemons, S. M., Reines, A. E., Plotkin, R. M., Gallo, E., & Greene, J. E. 2015, *The Astrophysical Journal*, 805, 12, doi: [10.1088/0004-637x/805/1/12](https://doi.org/10.1088/0004-637x/805/1/12)
- MacLeod, C. L., Brooks, K., Ivezić, Ž., et al. 2011, *The Astrophysical Journal*, 728, 26, doi: [10.1088/0004-637x/728/1/26](https://doi.org/10.1088/0004-637x/728/1/26)
- Massey, F. J. J. 1951, *Journal of the American Statistical Association*, 46, 68, doi: [10.1080/01621459.1951.10500769](https://doi.org/10.1080/01621459.1951.10500769)
- Mendez, A. J., Coil, A. L., Aird, J., et al. 2013, 770, 40, doi: [10.1088/0004-637x/770/1/40](https://doi.org/10.1088/0004-637x/770/1/40)
- Merloni, A., Heinz, S., & di Matteo, T. 2003, *MNRAS*, 345, 1057, doi: [10.1046/j.1365-2966.2003.07017.x](https://doi.org/10.1046/j.1365-2966.2003.07017.x)
- Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, *A&A*, 507, 1793, doi: [10.1051/0004-6361/200912497](https://doi.org/10.1051/0004-6361/200912497)
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, 12, 2825
- Plotkin, R. M., Gallo, E., Haardt, F., et al. 2016, *ApJ*, 825, 139, doi: [10.3847/0004-637X/825/2/139](https://doi.org/10.3847/0004-637X/825/2/139)
- Ramos Padilla, A. F., Wang, L., Malek, K., Efstathiou, A., & Yang, G. 2021, *Monthly Notices of the Royal Astronomical Society*, 510, 687, doi: [10.1093/mnras/stab3486](https://doi.org/10.1093/mnras/stab3486)
- Ranalli, P., Comastri, A., & Setti, G. 2003, *A&A*, 399, 39, doi: [10.1051/0004-6361:20021600](https://doi.org/10.1051/0004-6361:20021600)
- Reines, A. E., Greene, J. E., & Geha, M. 2013, *Astrophys. J.*, 775, 116, doi: [10.1088/0004-637X/775/2/116](https://doi.org/10.1088/0004-637X/775/2/116)
- Rodighiero, G., Daddi, E., Baronchelli, I., et al. 2011, *The Astrophysical Journal Letters*, 739, L40, doi: [10.1088/2041-8205/739/2/L40](https://doi.org/10.1088/2041-8205/739/2/L40)
- Salim, S., Boquien, M., & Lee, J. C. 2018, *ApJ*, 859, 11, doi: [10.3847/1538-4357/aabf3c](https://doi.org/10.3847/1538-4357/aabf3c)
- Salim, S., Lee, J. C., Janowiecki, S., et al. 2016, *ApJS*, 227, 2, doi: [10.3847/0067-0049/227/1/2](https://doi.org/10.3847/0067-0049/227/1/2)
- Sartori, L. F., Schawinski, K., Treister, E., et al. 2015, *MNRAS*, 454, 3722, doi: [10.1093/mnras/stv2238](https://doi.org/10.1093/mnras/stv2238)
- Satyalpal, S., Abel, N. P., & Secrest, N. J. 2018, *ApJ*, 858, 38, doi: [10.3847/1538-4357/aab7f8](https://doi.org/10.3847/1538-4357/aab7f8)
- Schawinski, K., Virani, S., Simmons, B., et al. 2009, *ApJL*, 692, L19, doi: [10.1088/0004-637X/692/1/L19](https://doi.org/10.1088/0004-637X/692/1/L19)
- Schmidt, K. B., Marshall, P. J., Rix, H.-W., et al. 2010, *ApJ*, 714, 1194, doi: [10.1088/0004-637X/714/2/1194](https://doi.org/10.1088/0004-637X/714/2/1194)
- Shirazi, M., & Brinchmann, J. 2012, *Monthly Notices of the Royal Astronomical Society*, 421, 1043, doi: [10.1111/j.1365-2966.2012.20439.x](https://doi.org/10.1111/j.1365-2966.2012.20439.x)
- Silverman, J. D., Mainieri, V., Lehmer, B. D., et al. 2008, *The Astrophysical Journal*, 675, 1025, doi: [10.1086/527283](https://doi.org/10.1086/527283)
- Spergel, D. N., Bean, R., Doré, O., et al. 2007, *ApJS*, 170, 377, doi: [10.1086/513700](https://doi.org/10.1086/513700)
- Stalevski, M. 2012, *Bulgarian Astronomical Journal*, 18, 3
- Stalevski, M., Fritz, J., Baes, M., Nakos, T., & Popović, L. Č. 2012, *Monthly Notices of the Royal Astronomical Society*, 420, 2756, doi: [10.1111/j.1365-2966.2011.19775.x](https://doi.org/10.1111/j.1365-2966.2011.19775.x)
- Stalevski, M., Ricci, C., Ueda, Y., et al. 2016, *Monthly Notices of the Royal Astronomical Society*, 458, 2288, doi: [10.1093/mnras/stw444](https://doi.org/10.1093/mnras/stw444)
- Stern, D., Assef, R. J., Benford, D. J., et al. 2012, *ApJ*, 753, 30, doi: [10.1088/0004-637X/753/1/30](https://doi.org/10.1088/0004-637X/753/1/30)
- Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, *AJ*, 124, 1810, doi: [10.1086/342343](https://doi.org/10.1086/342343)
- Ulrich, M.-H., Maraschi, L., & Urry, C. M. 1997, *Annual Review of Astronomy and Astrophysics*, 35, 445, doi: [10.1146/annurev.astro.35.1.445](https://doi.org/10.1146/annurev.astro.35.1.445)
- van der Maaten, L., & Hinton, G. 2008, *Journal of Machine Learning Research*, 9, 2579, <http://jmlr.org/papers/v9/vandemaaten08a.html>
- Wang, Z. F., Burke, C. J., Liu, X., & Shen, Y. 2023, *Monthly Notices of the Royal Astronomical Society*, 521, 99, doi: [10.1093/mnras/stad532](https://doi.org/10.1093/mnras/stad532)
- Wasleske, E. J., & Baldassare, V. F. 2023, *The Astronomical Journal*, 166, 64, doi: [10.3847/1538-3881/ace16b](https://doi.org/10.3847/1538-3881/ace16b)
- . 2024, *The Astrophysical Journal*, 971, 68, doi: [10.3847/1538-4357/ad5442](https://doi.org/10.3847/1538-4357/ad5442)

Wasleske, E. J., Baldassare, V. F., & Carroll, C. M. 2022, The Astrophysical Journal, 933, 37, doi: [10.3847/1538-4357/ac715b](https://doi.org/10.3847/1538-4357/ac715b)

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868, doi: [10.1088/0004-6256/140/6/1868](https://doi.org/10.1088/0004-6256/140/6/1868)

—. 2019, AllWISE Source Catalog, NASA IPAC DataSet, IRSA1, doi: [10.26131/IRSA1](https://doi.org/10.26131/IRSA1)

Yang, G., Boquien, M., Brandt, W. N., et al. 2022, ApJ, 927, 192, doi: [10.3847/1538-4357/ac4971](https://doi.org/10.3847/1538-4357/ac4971)

APPENDIX

A. PARAMETER DISTRIBUTIONS OF CLUSTERS

Below are shown the parameter distributions for the other eight clusters mentioned in Section 3.2.

B. PERCENTILE PLOTS OF EACH PARAMETER

Below we give the remaining percentile plots to show the shape of each clusters distributions within the t-SNE input parameters.

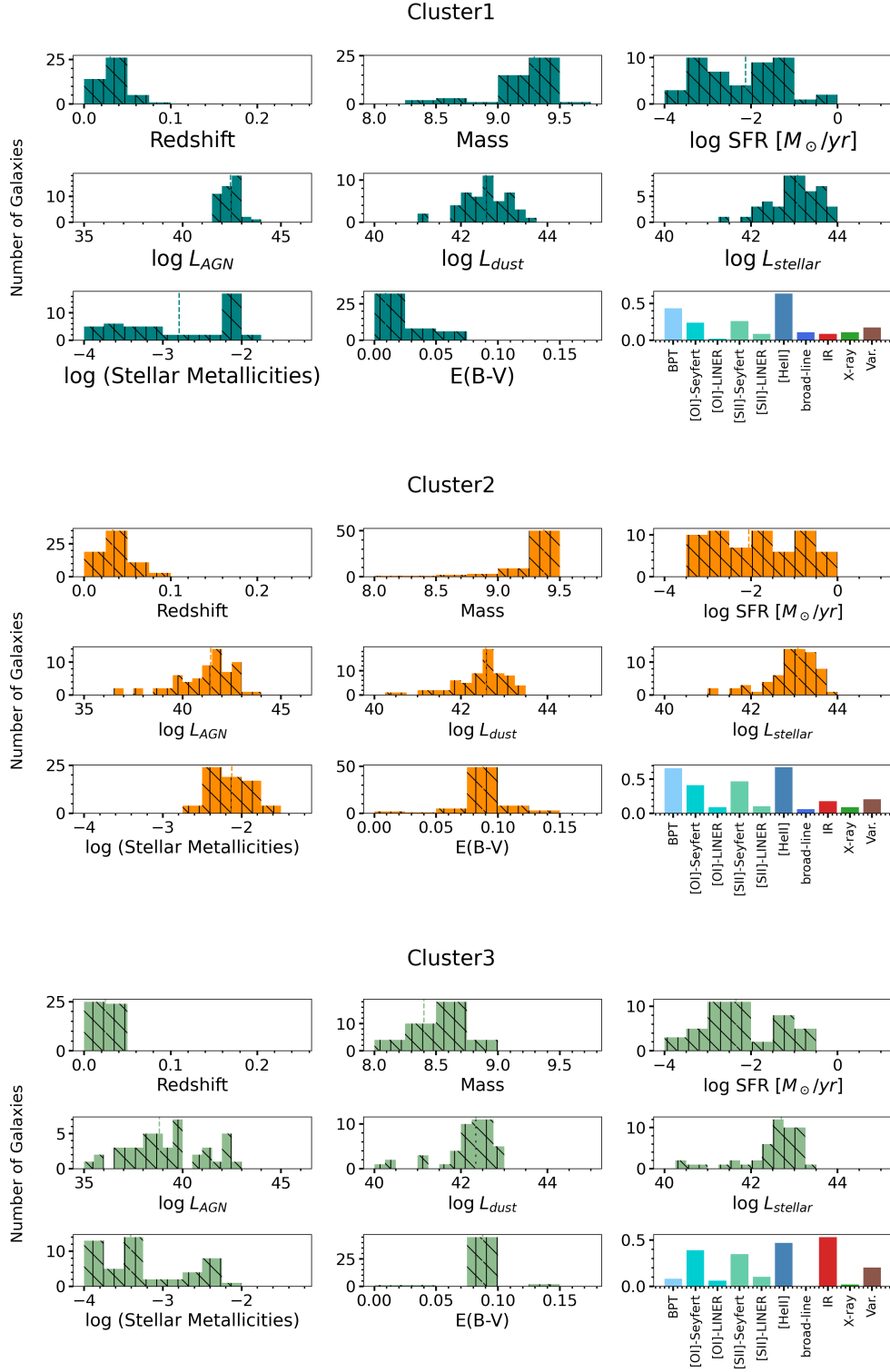


Figure 14. Distribution of host galaxy parameters for Clusters 1, 2, and 3 from the t-SNE dimensionality reduction.

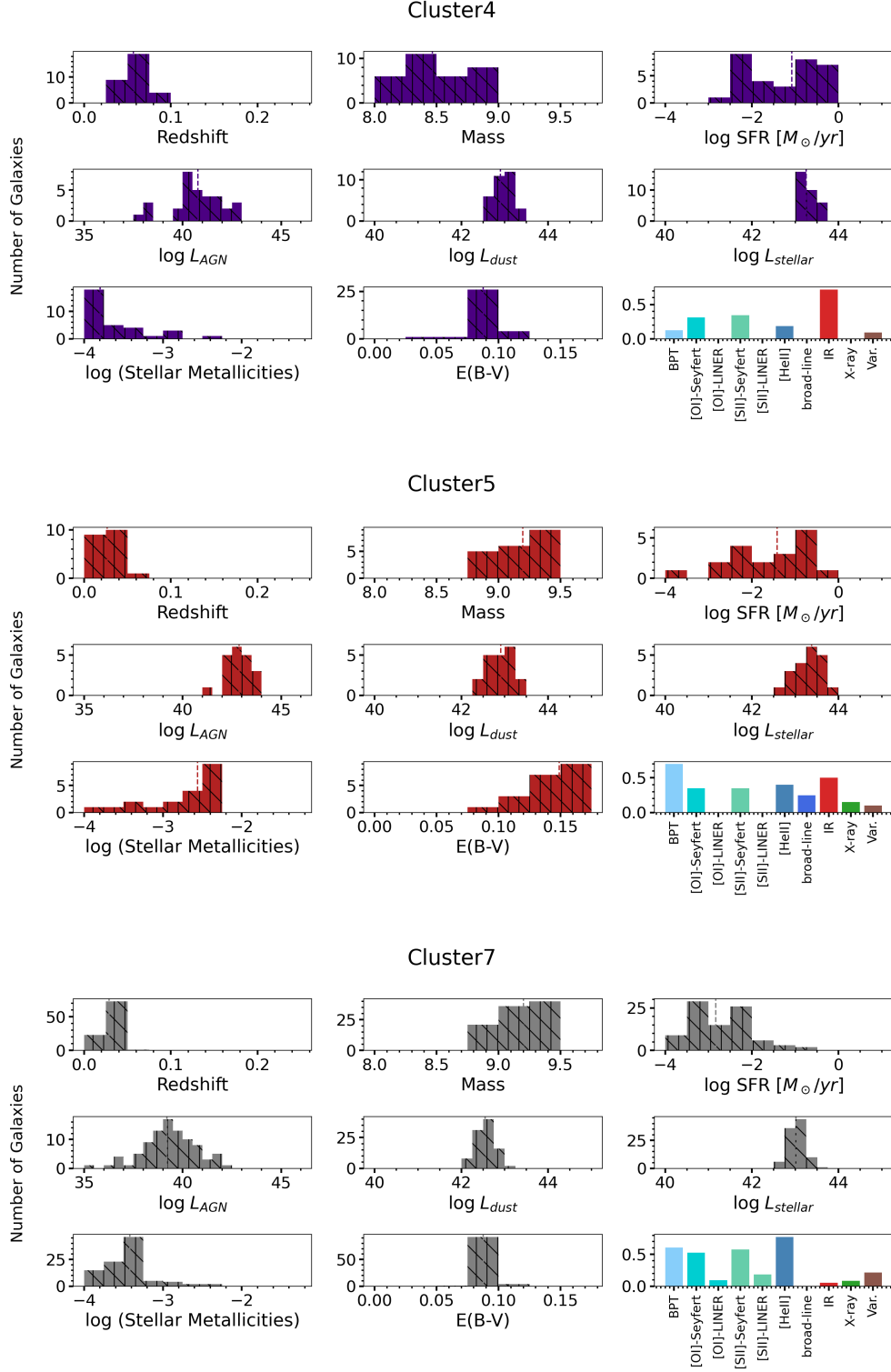


Figure 15. Distribution of host galaxy parameters for Clusters 4, 5, and 7 from the t-SNE dimensionality reduction.

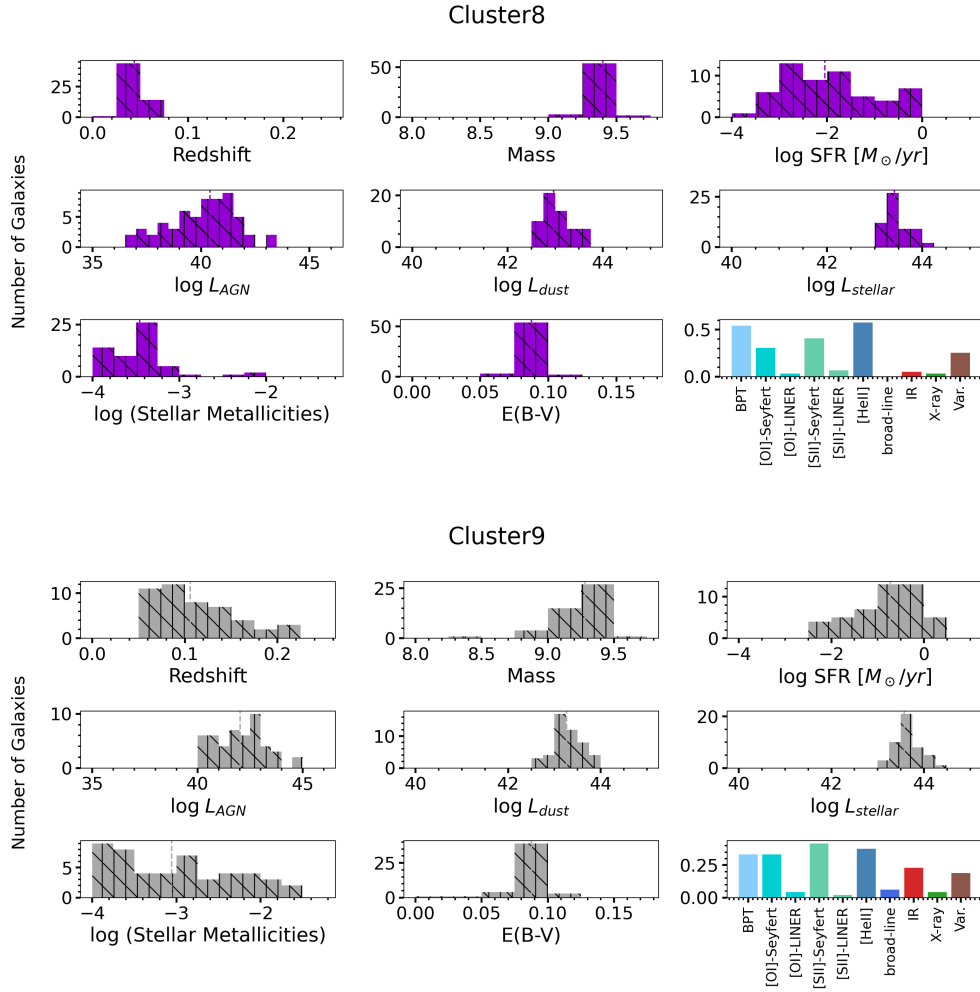


Figure 16. Distribution of host galaxy parameters for Clusters 8 and 9 from the t-SNE dimensionality reduction.

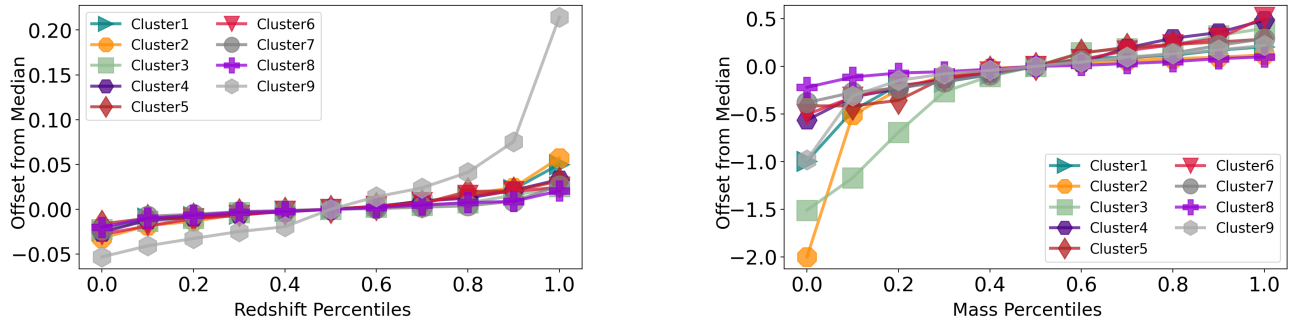


Figure 17. Percentile of the distribution vs. offset from median for each cluster within the t-SNE analysis for redshift and mass.

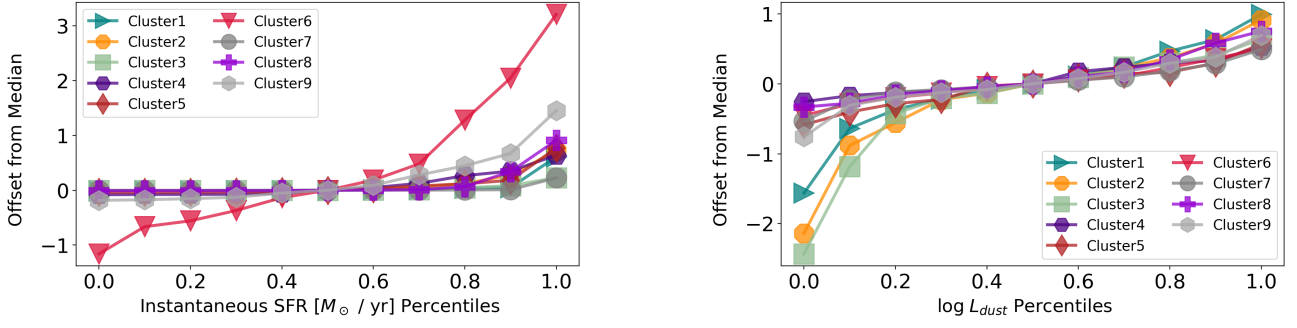


Figure 18. Percentile of the distribution vs. offset from median for each cluster within the t-SNE analysis for SFR and L_{dust} .

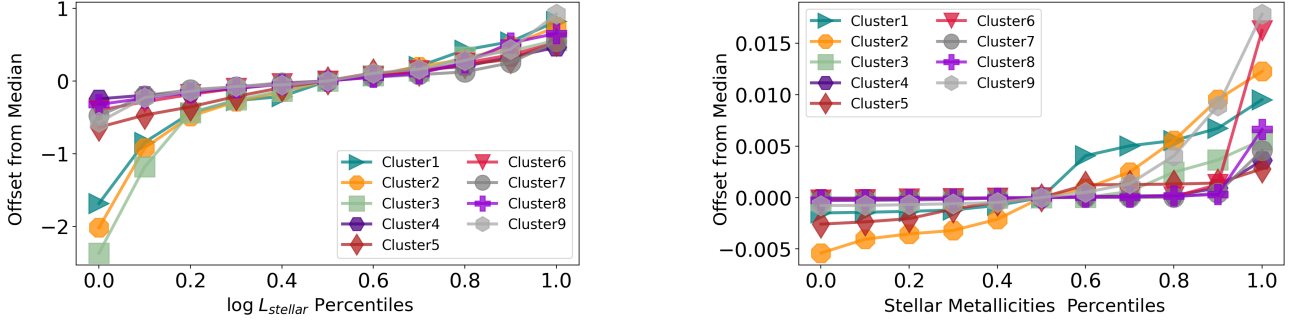


Figure 19. Percentile of the distribution vs. offset from median for each cluster within the t-SNE analysis for L_{stellar} and stellar metallicities.

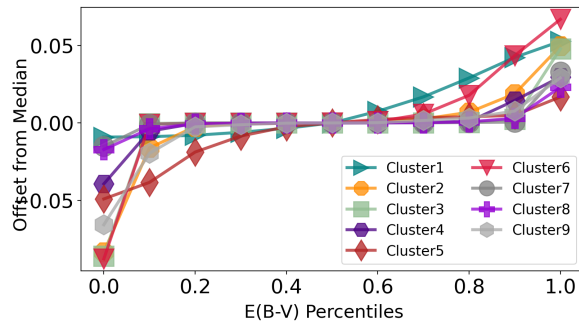


Figure 20. Percentile of the distribution vs. offset from median for each cluster within the t-SNE analysis for excess color.