

## Identifying Doppelgänger Active Galactic Nuclei across redshifts from spectroscopic surveys

SHREYA SAREEN<sup>1,2,\*</sup> AND SWAYAMTRUPA PANDA<sup>3,†</sup>

<sup>1</sup> University of California, Berkeley, 110 Sproul Hall, Berkeley, CA 94720, United States

<sup>2</sup> Lockheed Martin Rotary Mission Systems, New York, United States

<sup>3</sup> International Gemini Observatory/NSF NOIRLab, Casilla 603, La Serena, Chile

### ABSTRACT

Active Galactic Nuclei (AGNs) are among the most luminous objects in the universe, making them valuable probes for studying galaxy evolution. However, understanding how AGN properties evolve over cosmic time remains a fundamental challenge. This study investigates whether AGNs at low redshift (nearby) can serve as proxies for their high-redshift (distant) counterparts by identifying spectral “doppelgängers”, AGNs with remarkably similar emission line properties despite being separated by vast cosmic distances. We analyze key spectral features of bona fide AGNs using the Sloan Digital Sky Survey’s Data Release 16, including continuum and emission lines: Nitrogen (N V), Carbon (C IV), Magnesium (Mg II), Hydrogen-beta (H $\beta$ ), and Iron (Fe II - optical and UV) emission lines. We incorporated properties such as equivalent width, velocity dispersion in the form of full width at half maximum (FWHM), and continuum luminosities (1350Å, 3000Å, and 5100Å) closest to these prominent lines. Our initial findings suggest the existence of multiple AGNs with highly similar spectra, hinting at the possibility that local AGNs may indeed share intrinsic properties with high-redshift ones. We showcase here one of the better candidate pairs of AGNs resulting from our analyses.

### 1. INTRODUCTION

Active Galactic Nuclei are compact regions at the centers of galaxies where supermassive black holes accrete matter, releasing vast amounts of energy across the electromagnetic spectrum (Antonucci 1993; Padovani et al. 2017). Their brightness makes them visible across cosmic distances, allowing astronomers to probe the conditions of the early universe. Despite extensive studies, the evolution of AGN properties—such as black hole accretion rates, gas composition, and feedback mechanisms—remains a subject of ongoing debate. Understanding these changes is crucial for building a comprehensive picture of galaxy formation and the co-evolution of galaxies and their central black holes (Kormendy & Ho 2013; Zavala et al. 2024).

A promising yet underexplored method of studying AGN evolution is through spectral comparison between AGNs at different redshifts. This project focuses on identifying AGN “doppelgängers” which are pairs of

AGNs that exhibit nearly identical emission line characteristics but exist at vastly different epochs in the universe’s history. In the context of AGNs, a doppelgänger refers to a galaxy or quasar that closely resembles another AGN in observable features, such as line profiles and intensities, while potentially differing in intrinsic physical conditions. If such spectral twins can be found, especially across low and high-redshift populations, they could serve as a bridge for inferring the properties of early-universe AGNs using well-resolved, nearby counterparts.

A typical AGN spectrum contains three primary components: an underlying ionizing continuum from the AGN itself, absorption lines from intervening material, and both broad and narrow emission lines originating from gas in the vicinity of the black hole. The combined information from these spectral features offers critical insights into the dynamics, energetics, and composition of the AGN’s immediate environment (Alexander & Hickox 2012; Netzer 2015).

To explore this, we analyze key spectral features—specifically the Mg II, H $\beta$ , C IV, and Fe II emission (optical and UV complexes) lines. By comparing parameters like equivalent width, velocity dispersion in the form of full width at half maximum (FWHM), and continuum luminosities (1350Å, 3000Å, and 5100Å)

Corresponding author: Swayamtrupta Panda  
swayamtrupta.panda@noirlab.edu

\* NOIRLab REU intern

† Gemini Science Fellow

closest to these prominent lines, we aim to identify pairs of AGNs with compelling spectral similarity. The presence of such doppelgängers would suggest a degree of universality in AGN behavior, supporting the use of local AGNs as proxies for those at high redshift and validating the use of local scaling relations (Vestergaard & Peterson 2006; Bentz et al. 2013) to infer BH properties at higher redshifts. However, if no such matches are found, it would imply significant evolution in AGN physics, highlighting the need to treat AGNs at different epochs as fundamentally distinct.

## 2. METHODOLOGY

To investigate the possibility of spectral “doppelgängers” among AGNs, this study involves a comparative analysis of AGN spectra across different redshifts using the Sloan Digital Sky Survey (SDSS) Data Release 16<sup>1</sup> (Wu & Shen 2022), which contains 750,414 AGNs. We focus on identifying AGNs that exhibit similar emission-line characteristics despite being separated by vast cosmic distances. The analysis centers on key spectral features critical to understanding AGN physics: Magnesium (Mg II) at 2800Å, Hydrogen (H $\beta$ ) at 4861Å, Carbon (C IV) at 1549Å, Nitrogen (N V) at 1240Å, and Iron (Fe II) emission in both UV (2700–2900 Å) and optical (4434–4684 Å) regimes. In addition, we examine continuum luminosities from the accretion disk at 1350 Å, 3000 Å, and 5100 Å. These features are analyzed to extract, for the moment, equivalent widths (which quantify the strength of these emission lines) and full width at half maximum (FWHM) for velocity dispersion.

### 2.1. low- and high-z SDSS sample

*Are the doppelgängers sky region dependent? Are we finding more doppelgängers in specific regions of the sky?* – To investigate potential doppelgängers across cosmic time, the dataset was divided into two redshift-based subsets inspired by a previous work (Panda & Skorek 2022) that performed spectral analyses based on an earlier QSO data release (Rakshit et al. 2020). The low-redshift sample spans a redshift range of 0.617 to 0.89 and includes 57,268 AGNs characterized by emission lines such as Fe II (UV and optical), Mg II, and H $\beta$ . The high-redshift sample covers redshifts from 2.09 to 2.25 and consists of 51,927 AGNs, prominently featuring lines like N V, C IV, Fe II (UV), and Mg II. These divisions ensure that comparable spectral features fall within the observable wavelength ranges of each subset (see Figure 1). An important question addressed during

this analysis is whether doppelgänger AGNs are region-dependent—that is, whether they tend to occur more frequently in specific areas of the sky. Our findings show that matched AGN pairs are distributed across diverse sky regions, suggesting that their spectral similarities are not a product of localized sky clustering or observational biases, but rather reflect broader, potentially universal physical processes.

### 2.2. Doppelganger detection pipeline

To identify spectral doppelgängers, an end-to-end detection pipeline was developed, as shown with the flowchart in Figure 1. The process begins with data cleaning, including the removal of unreliable or invalid measurements and the selection of relevant spectral features. Following this, dimensionality reduction techniques are applied to condense the high-dimensional feature space while retaining critical information. Principal Component Analysis<sup>2</sup> (PCA, Jolliffe 2011) was employed to capture the major sources of variance in the dataset, allowing for a compact and interpretable representation of spectral characteristics. This transformation enabled efficient similarity searches using a nearest-neighbors algorithm based on Euclidean distances in the PCA’s hyperparameter space, which systematically compared each AGN to others within the reduced space to identify candidate pairs with the smallest separation. This pipeline was applied independently to both low- and high-redshift samples, and later to the merged dataset, allowing for the identification of cross-redshift AGN pairs that share similar spectral profiles.

## 3. RESULTS AND DISCUSSIONS

### 3.0.1. PCA

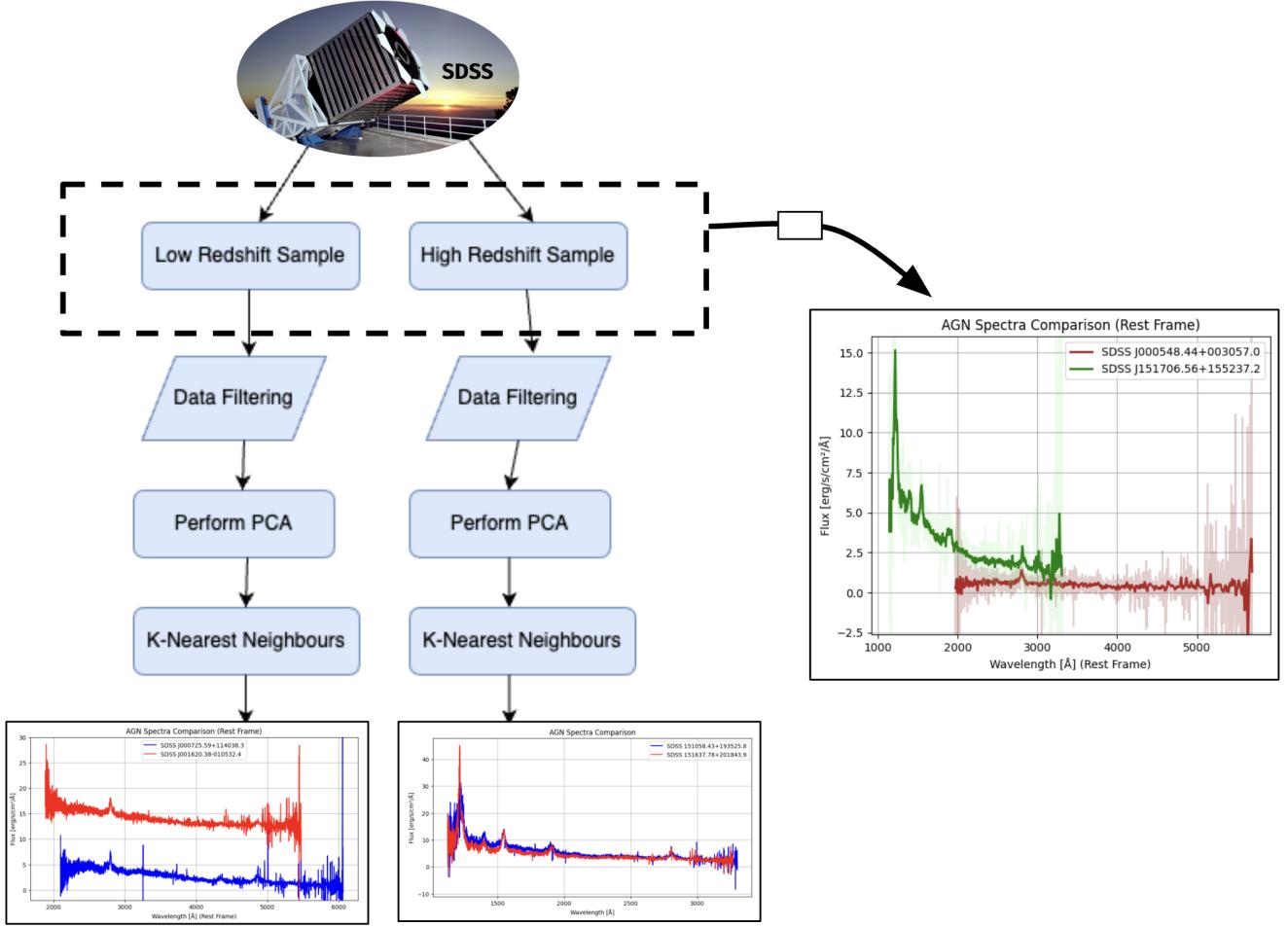
PCA was used to transform the original spectral features into a lower-dimensional space while retaining the maximum variance. For the high-z sample, four principal components captured 100 % of the spectral variance (PC1 = 40.9 %, PC2 = 23.1 %, PC3 = 19.6 %, PC4 = 16.4 %), while for the low-z sample, PC1 and PC2 alone accounted for the total variance (PC1 = 76.5 %, PC2 = 23.5 %). This enabled 2D visualization and clustering of spectrally similar AGNs.

### 3.0.2. Nearest-neighbors algorithm and tweaks

Once dimensionality was reduced, a nearest-neighbors algorithm was applied to locate AGN pairs with minimal spectral distance in the transformed space. This allowed the identification of  $\sim$ 600 potential doppelgänger pairs.

<sup>1</sup> [http://quasar.astro.illinois.edu/paper\\_data/DR16Q/](http://quasar.astro.illinois.edu/paper_data/DR16Q/)

<sup>2</sup> <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>



**Figure 1.** Flowchart for an end-to-end pipeline to retrieve doppelgängers across redshift using SDSS DR16 Spectroscopic Database (Wu & Shen 2022). On the bottom left panels, we are highlighting pairs of AGN spectra for the low-redshift sample and the high-redshift sample. On the extreme right, we highlight one pair of AGN spectra resulting from the analysis for the merged sample. We note that the detection of these doppelgängers is independent of the sky region.

Adjustments to the number of neighbors and distance metrics were made iteratively to balance sensitivity and specificity, ensuring that identified matches had high-quality, interpretable spectral similarities.

When low- and high-z samples were combined, PCA revealed that the first two components ( $PC1 = 77.1\%$ ,  $PC2 = 22.9\%$ ) were sufficient to represent 100 % of the spectral variance. Analysis of this merged space identified multiple AGN pairs—each pair consisting of one low-z and one high-z AGN—with strikingly similar spectral features. Emission lines such as Mg II, H $\beta$ , Fe II, and C IV aligned well between these matched pairs, including consistency in equivalent widths, FWHM, and line ratios. The similarities, despite differences in redshift and sky position, suggest intrinsic parallels in AGN physics that transcend cosmic time.

Future work will focus on expanding the scope and depth of the analysis. One key direction is the incorporation of newly released spectroscopic data from the Dark Energy Spectroscopic Instrument (DESI, DESI Collaboration et al. 2025), which offers a significantly larger sample of AGNs across a broad redshift range ( $z = 0\text{--}4$ ). Additionally, the James Webb Space Telescope (JWST, Gardner et al. 2006) will be leveraged for its high-resolution infrared capabilities, particularly to observe faint, early-universe AGNs at redshifts beyond  $z \gtrsim 4$ . Lastly, we aim to refine the doppelgänger detection pipeline using advanced non-linear dimensionality reduction methods and supervised learning techniques to improve the precision and reliability of AGN matching across epochs.

#### 4. SUMMARY AND FUTURE WORK

This research was conducted as part of the “Research Experience in Optical-Infrared Astronomy at NOIRLab in Chile” REU Site funded by NSF Award Number 2349023. SP is supported by the International Gemini Observatory, a program of NSF NOIRLab, which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the U.S. National Science Foundation, on behalf of the Gemini partnership of Argentina, Brazil, Canada, Chile, the Republic of Korea, and the United States of America.

*Facilities:* SDSS

*Software:* numpy (Oliphant 2015), matplotlib (Hunter 2007), scipy (Virtanen et al. 2020), astropy (Astropy Collaboration et al. 2013, 2018)

## REFERENCES

- Alexander, D. M., & Hickox, R. C. 2012, NewAR, 56, 93, doi: [10.1016/j.newar.2011.11.003](https://doi.org/10.1016/j.newar.2011.11.003)
- 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)
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: [10.1051/0004-6361/201322068](https://doi.org/10.1051/0004-6361/201322068)
- 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)
- Bentz, M. C., Denney, K. D., Grier, C. J., et al. 2013, ApJ, 767, 149, doi: [10.1088/0004-637X/767/2/149](https://doi.org/10.1088/0004-637X/767/2/149)
- DESI Collaboration, Abdul-Karim, M., Adame, A. G., et al. 2025, arXiv e-prints, arXiv:2503.14745, doi: [10.48550/arXiv.2503.14745](https://doi.org/10.48550/arXiv.2503.14745)
- Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, SSRv, 123, 485, doi: [10.1007/s11214-006-8315-7](https://doi.org/10.1007/s11214-006-8315-7)
- Hunter, J. D. 2007, Computing in Science and Engineering, 9, 90, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55)
- Jolliffe, I. 2011, Principal Component Analysis, ed. M. Lovric (Berlin, Heidelberg: Springer Berlin Heidelberg), 1094–1096, doi: [10.1007/978-3-642-04898-2\\_455](https://doi.org/10.1007/978-3-642-04898-2_455)
- 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)
- Netzer, H. 2015, ARA&A, 53, 365, doi: [10.1146/annurev-astro-082214-122302](https://doi.org/10.1146/annurev-astro-082214-122302)
- Oliphant, T. 2015, NumPy: A guide to NumPy, 2nd edn., USA: CreateSpace Independent Publishing Platform. <http://www.numpy.org/>
- Padovani, P., Alexander, D. M., Assef, R. J., et al. 2017, A&A Rv, 25, 2, doi: [10.1007/s00159-017-0102-9](https://doi.org/10.1007/s00159-017-0102-9)
- Panda, S., & Skorek, E. J. 2022, in XL Polish Astronomical Society Meeting, ed. E. Szuszkiewicz, A. Majczyna, K. Małek, M. Ratajczak, E. Niemczura, U. Bak-Stęślicka, R. Poleski, M. Bilicki, & Ł. Wyrzykowski, Vol. 12, 72–75, doi: [10.48550/arXiv.2111.14516](https://doi.org/10.48550/arXiv.2111.14516)
- Rakshit, S., Stalin, C. S., & Kotilainen, J. 2020, ApJS, 249, 17, doi: [10.3847/1538-4365/ab99c5](https://doi.org/10.3847/1538-4365/ab99c5)
- Vestergaard, M., & Peterson, B. M. 2006, ApJ, 641, 689, doi: [10.1086/500572](https://doi.org/10.1086/500572)
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)
- Wu, Q., & Shen, Y. 2022, ApJS, 263, 42, doi: [10.3847/1538-4365/ac9ead](https://doi.org/10.3847/1538-4365/ac9ead)
- Zavala, J. A., Bakx, T., Mitsuhashi, I., et al. 2024, ApJL, 977, L9, doi: [10.3847/2041-8213/ad8f38](https://doi.org/10.3847/2041-8213/ad8f38)