Carnegie Observatories Astrophysics Series, Vol. 1:
Coevolution of Black Holes and Galaxies, 2004
ed. L. C. Ho (Pasadena: Carnegie Observatories,
http://www.ociw.edu/ociw/symposia/series/symposium1/proceedings.html)

The AGN Population and Luminosity Function in the Sloan Digital Sky Survey

L. HAO and M. A. STRAUSS
Princeton University

Abstract

Low-redshift AGNs are identified from the spectroscopic data of the Sloan Digital Sky Survey (SDSS) based on their emission line features. We have applied careful stellar subtraction to every galaxy spectrum before making emission line measurements. The distribution of the Hα FWHMs of the emission line galaxies is bimodal with two populations separated at about 1200km/s. Therefore the galaxies having FWHM(Hα)$>$1200km/s are classified as broad-line AGNs. The narrow-line AGNs are identified using emission line ratio diagnostic diagrams. 1279 broad-line and 2694 narrow-line AGNs are identified from over 80,000 galaxy spectra complete in 1151 square degrees. We calculated the detection probability function of each AGN using Monte-Carlo simulations, assuming the nuclear luminosity is independent of the host galaxy luminosity, and we obtain the Hα luminosity function for both Seyfert 1s and Seyfert 2s. The relative ratio of Seyfert 1s to Seyfert 2s is about 1 at low luminosities. However toward high luminosities, the number of Seyfert 2s drops off very quickly and Seyfert 1s greatly outnumber Seyfert 2s. Detailed tests of the luminosity function are conducted. We show strong evidence that the nuclear luminosity is not strongly correlated with the host galaxy luminosity.

1.1 Introduction

Active Galactic Nuclei (AGNs) can be identified by the characteristics of their prominent emission lines. Some AGNs, such as Seyfert 1s and quasars have very prominent broad permitted emission lines. Other AGNs such as Seyfert 2s only show narrow emission lines. The emission line types in narrow-line AGNs are very similar to those in normal starforming galaxies, except that they appear to represent a greater ionization range. For example, both the high ionization lines such as [NeIII], [NeIV], [OIII] and the low ionization lines, such as [OI] and [NI] are stronger than those of starforming galaxies. Based on these characteristics, broad-line AGNs can be identified by checking the FWHM of the permitted emission lines, while narrow-line AGNs can be identified by their emission line ratios. In particular, emission line ratio diagrams have been developed in separating narrow-line AGNs from the normal star-forming galaxies. The line ratio diagram was first introduced by Baldwin, Phillips and Terlevich (1982), and later revised by Veilleux & Osterbrock (1987). By using the combination of line ratios: [OIII]λ5007/Hβ, [NII]λ6584/Hα, [OI]λ6300/Hα and [SII]$\lambda\lambda 6817, 23$/Hα, the narrow-line AGNs and starforming galaxies are well separated by their locations in the diagrams. Recently, Kewley et al. (2001) re-analyzed the galaxies' positions in the diagnostic diagrams using theoretical models for starburst galaxies, and developed theoretical separations between AGNs and starforming galaxies in the diagnostic diagrams.

In this paper, we select AGNs by their emission line features from the Sloan Digital Sky Survey (SDSS) data. The SDSS (York et al. 2000, see also Gunn et al. 1998) is an imaging and spectroscopic survey that will eventually cover approximately one-quarter of the celestial sphere and collect $\sim 10^6$ galaxy spectra. The SDSS spectra are taken with two fiber-fed spectrographs, covering the wavelength range 3800-9200 over 4096 pixels. The resolution $\lambda/\Delta\lambda$ varies between 1850 and 2200. The huge number and high quality of the spectra available in the SDSS make it very valuable in the study of selecting AGNs based on their emission line characteristics.

1.2 Data Sample

The sample of galaxies analyzed here is complete over 1151 square degrees of sky and includes all objects that are spectroscopically classified as galaxies with Petrosian r band magnitudes less than 17.77 (called galaxy target objects, see Strauss et al. 2002). These galaxies are low redshift galaxies with a median redshift at about 0.1 (Figure 1.5). About 200 galaxies ($\sim 0.25\%$ of the sample) have $z \geq 0.33$ and are ignored since their Hα emission lines would be redshifted out of the SDSS observational range. The final galaxy sample includes 83,839 galaxies. AGNs selected from this sample are thus low redshift and most of them are Seyferts. Therefore, in this paper the name AGNs and Seyferts are used interchangeably.

1.3 Stellar Subtraction

The accurate measurement of emission lines is very important in properly identifying AGNs. However, the nuclear emission lines are very susceptible to contamination by the stellar absorption lines coming from the host galaxy. Therefore before the AGN selection, we need to remove the underlying stellar absorption lines. The basic idea is to use a library of absorption-line templates to model the non-emission-line region of the spectrum. Usually the template library is composed of either star spectra (Saraiva et al. 2001; Kauffmann et al. 2004) or absorption-line galaxy spectra (Ho et al. 1997a). In this study we build the templates in a slightly different way. We apply Principal Component Analysis (PCA) to several hundred pure absorption-line galaxies and use the resultant eigenspectra as the absorption-line templates. PCA carries out the linear transformation among the galaxy spectra, and can extract the most prominent features from the sample and concentrate them into the first few eigenspectra. As a result we only need to use the first few eigenspectra as the templates, which makes the size of the absorption-line template library shrink from several hundred galaxies to a few. Furthermore, the eigenspectra are orthogonal, making the subtraction fitting using these templates unique.

The pure absorption-line galaxies are carefully picked from the SDSS data to not have weak emission line signature. This is done by checking the equivalent width of Hα and [OII]λ3727. Figure 1.1 shows the first six resultant eigenspectra. These eigenspectra capture the main features of the pure absorption-line galaxies (see the 1st eigenspectrum) as well as the major continuum variances (see e.g. 2nd and 3rd) and metallicity variances reflected as subtle metal lines (see e.g. 2nd, 4th, 5th) among the galaxies. Moreover higher eigenspectra

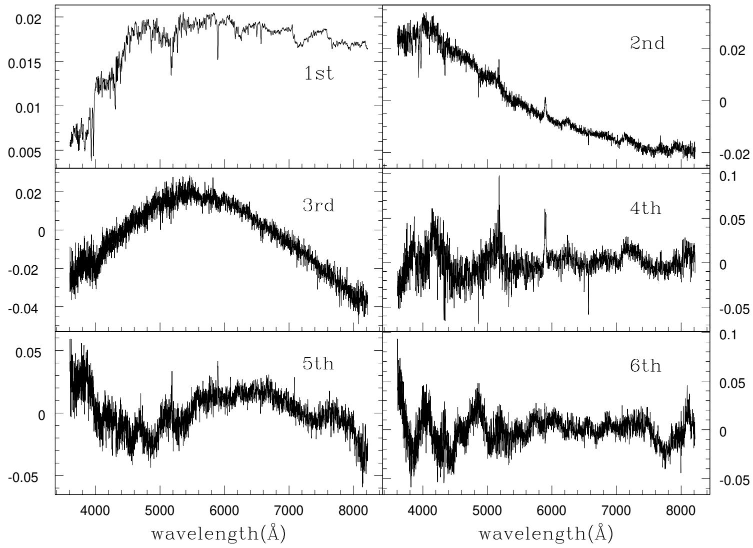

Fig. 1.1. The first six eigenspectra of the PCA of several hundred pure absorption line galaxies. These eigenspectra can be used as stellar templates in stellar subtraction.

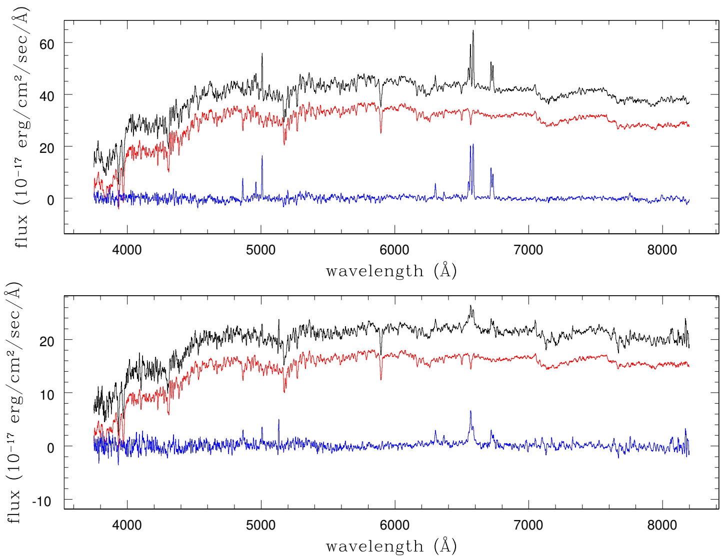

Fig. 1.2. Two examples of the performances of the stellar subtraction. In each panel, the black (top) is the original spectrum, red (middle) is the stellar spectrum built by fitting the emission line free regions with the 10 templates, and the blue (bottom) is the spectrum after the stellar subtraction. The stellar subtraction helps recover the submerged weak emission lines as well as correct the relative emission line ratios.

contain more noise and less signal. In this study, I will use the first 8 eigenspectra as the templates.

These absorption-line templates are not appropriate for the QSO spectra that don't include any stellar absorption features, and those galaxy spectra that include young stellar population. We thus added a simple power-law template, $e_{powerlaw} = C\lambda^{-1.5}$ and an A star spectrum to the template library. The addition of the two spectra greatly improves the behavior of the stellar/continuum subtraction for the two kinds of spectra. Figure 1.2 gives two examples of the stellar subtraction using the 8 eigenspectra + the power-law continuum + the A star spectrum as the stellar templates. As can be seen, weak emission lines such as Hβ that were originally submerged in the stellar absorption line are recovered and emission line ratios such as Hα/[NII] are corrected by the subtraction. Furthermore, after the stellar subtraction, weak emission lines such as [OI]λ6363, [NII]λ5200 can clearly be seen. The stellar subtraction is also very important in identifying the weak broad line AGNs as shown in the bottom panel, where the shape of the Hα line is corrected after the subtraction.

We apply the stellar subtraction to all the 83,839 galaxies in the data sample. The emission lines are measured with the subtracted spectrum assuming Gaussian profiles. To take into account of the possibility of some AGNs having both broad and narrow emission line components, we consider two models for Hα emission lines: one has only one Hα component; one has two Hα components. The χ^2 of the two model fittings are compared to decide which model is more suitable for data. This assures us that we won't miss any AGNs such as those classified as Seyfert 1.5s.

1.4 AGN Selection and Result

After the emission lines are measured, we can begin to select AGNs. Broad-line AGNs are easily identified by their broad permitted emission line widths. Figure 1.3 shows the distribution of the Hα FWHM for over 40,000 galaxies. If the Hα emission line fitting favors the model with two Hα components, the broader FWHM is plotted here. Interestingly, the distribution is bimodal, with two populations separating at about FWHM(Hα) $\approx$ 1,200 km/s. The huge peak of the narrow Hα group not shown on the plot reflects the fact that most of the galaxies which are normal starforming galaxies have narrow Hα emission lines. Naturally, we would classify the galaxies that have FWHM(Hα)$>$1,200km/s as broad-line AGNs. Based on this criterion, 1279 broad-line AGNs are selected.

The narrow-line AGNs can be identified using the traditional line ratio diagnostic diagrams. Due to the vast quantity of data available, we notice some structures of galaxies' distribution in the diagrams. Figure 1.4 gives the contour plot of over 40,000 galaxies in the diagnostic diagram. The solid lines are the AGN and starforming galaxy separation lines from the models of Kewley et al. (2001). As can be seen, the lines agree pretty well with the real density contour shape. We will use these lines as the selection criteria and choose those galaxies located at the upper-right of the lines in all of the three diagrams to be narrow-line AGNs. Based on this criterion, 2694 are identified as narrow-line AGNs.

Figure 1.5 gives the redshift and Petrosian magnitude distribution for identified broad and narrow-line AGNs compared with those of normal galaxies. From this plot it seems that there are more normal galaxies in low redshift than AGNs, and the galaxies that host AGNs are slightly more luminous than normal galaxies.

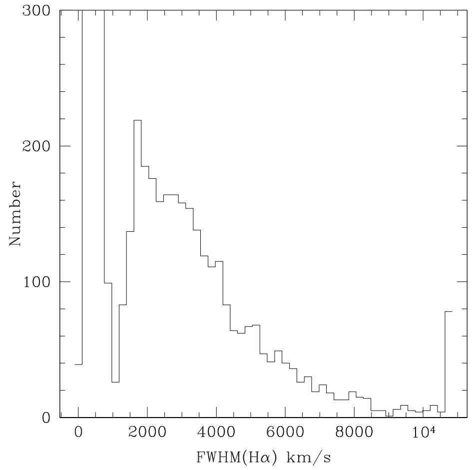

Fig. 1.3. The distribution of the $\mathrm{H}\alpha$ FWHM for over 40,000 galaxies. The distribution is bimodal. The first group peaks at about 200km/s, with the height of the peak in the order of 10^4. The second group with $\mathrm{H}\alpha$ FWHM $> 1{,}200$ km/s represents the broad-line AGNs.

1.5 Luminosity Function

1.5.1 Luminosity Function Definition

With the complete sample of AGNs selected, we measure the AGN luminosity function (LF). The AGN luminosity function $\Phi(L)$ is defined as the number of AGNs per unit volume per unit luminosity. We choose to use $\mathrm{H}\alpha$ luminosity as the variable parameter in the luminosity function, since it is closely connected with the nuclear activity. The non-parameterized LF is derived via the maximum likelihood method.

We define $p_i(L)$ as the detection probability function for the ith AGN in the sample (detailed discussion in § 1.5.2). The likelihood that this AGN has luminosity L is:

$$\mathcal{L}_i dL = \frac{p_i(L)\Phi(L)dL}{\int_0^\infty \Phi(L')p_i(L')dL'}$$

The maximum likelihood solution $\Phi(L)$ is the function that maximizes the overall likelihood $\mathcal{L} = \prod_i \mathcal{L}_i$. Therefore $\Phi(L)$ is obtained by minimizing $\mathcal{S} = -2\ln\mathcal{L}$. The detailed formulae and error estimation can be found in Efstathiou, Ellis & Peterson 1988, Koranyi & strauss 1997 and Blanton 2000. In this paper the luminosity functions for narrow-line and broad-line AGNs are evaluated separately. The former one considers all narrow-line AGNs as well as the narrow components of those broad-line AGNs whose narrow components themselves are AGNs by their line rations. For the broad-line AGN luminosity function, only the broad component of $\mathrm{H}\alpha$ emission line is used.

1.5.2 Detection Probability Function

As described in § 1.4, the AGN selection criterion for broad and narrow-line AGNs are based on their spectral characteristics. Correspondingly, their detection probabilities are very complicated to evaluate. The detection probability function depends on not only the nuclear luminosity, but also the host galaxy luminosity and the associated noise in the spec-

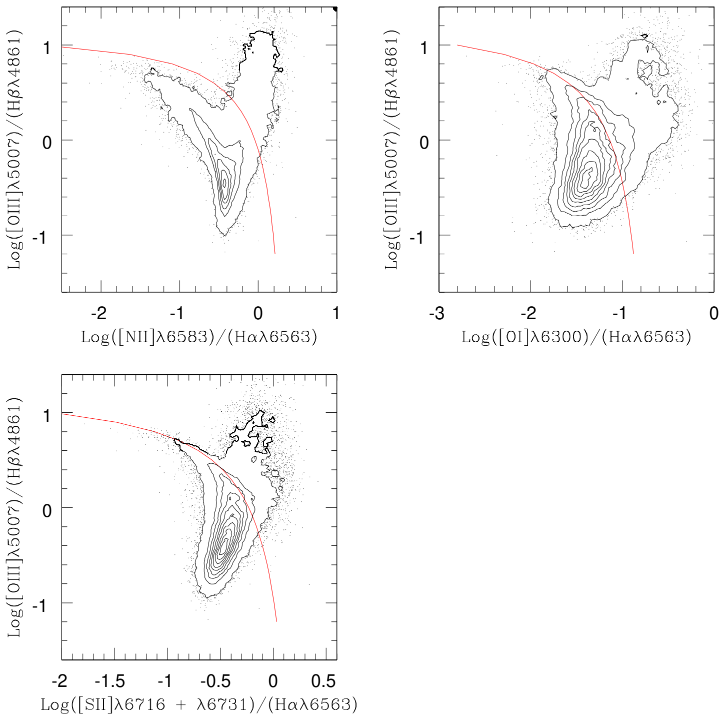

Fig. 1.4. The density contour plot of over 40,000 galaxies in the emission line ratio diagnostic diagrams. The solid lines are the theoretical separation lines of AGNs and starforming galaxies taken from Kewley et al. (2001).

trum. Therefore it is not possible to calculate an analytical form of the detection probability function, and we evaluate it through a Monte-Carlo simulation. Since AGNs are selected by their emission line features, we will use the pure emission line spectrum containing only the emission lines involved in the AGN selection to represent the nuclear spectrum. What is left represents the host galaxy spectrum. During the simulation, we make an assumption that the nuclear luminosity in each AGN is *independent* of the host galaxy luminosity. Based on this assumption, for each AGN we scale down the nuclear spectrum according to certain Hα luminosity L, but keep the host galaxy spectrum and its associated noise unchanged. Since the noise within the emission line region will decrease with the changed nuclear spectrum, a noise spectrum is randomly generated and added back to the spectrum to compensate for it. We apply our whole AGN selection procedure including the stellar subtraction to this new spectrum with weakened nuclei and check whether or not it still is recognized as an AGN. For each Hα luminosity, the same procedure is repeated several times, each with a noise spectrum randomly generated. The detection probability for this Hα luminosity is defined to be fraction of times the new spectrum is selected as an AGN. By varying the Hα luminosity L, we get the detection probability function $p_i(L)$ for this AGN. Furthermore, we make sure that the simulated galaxy with weakened nuclear luminosity still satisfy the SDSS spectroscopic selection criterion for galaxies. That is, the overall r band Petrosian magnitude at

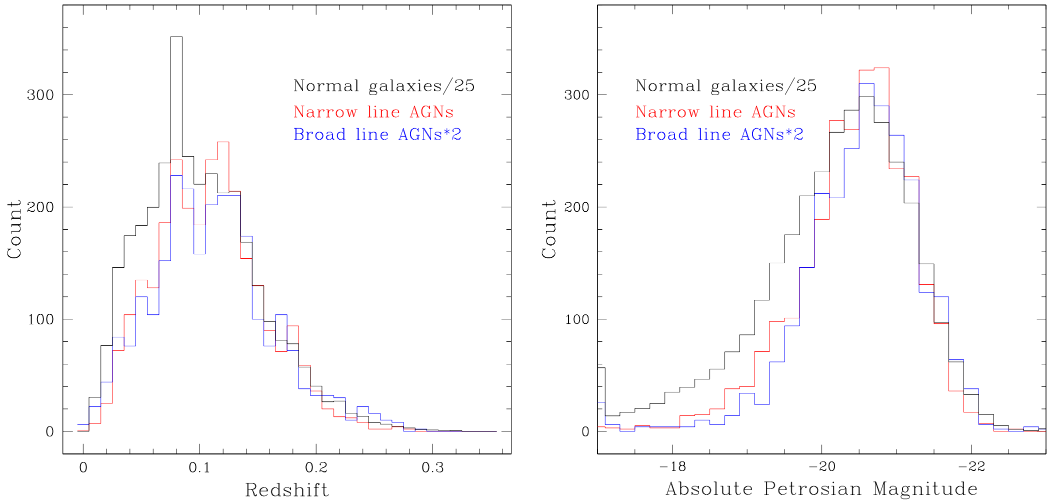

Fig. 1.5. The redshift and r band Petrosian absolute magnitude distribution for broad-line (blue) and narrow-line AGNs (red) compared with those of normal galaxies (black).

a certain Hα luminosity L is still less than 17.77. Otherwise, the detection probability for this AGN at this Hα luminosity L will be directly put to zero.

1.5.3 Luminosity Function Normalization and Results

With the detection probability function measured, we can evaluate the luminosity function using the maximum likelihood method (§ 1.5.1). However, the normalization of the luminosity function cannot be derived from this method and has to be evaluated separately. This is done by directly measuring the number density of AGNs, expressed as:

$$n = \frac{1}{V} \sum_i \frac{\int_0^\infty \Phi(L)dL}{\int_0^\infty \Phi(L)p_i(L)dL}$$

where V is the volume spanned by the galaxies in the sample and $\sum_i$ sums over all AGNs included in this LF calculation.

Figure 1.6 gives the luminosity functions for Seyfert 1s and Seyfert 2s separately. The luminosity functions cover a luminosity range of over 10^3 and are well-fit with a broken power-law, especially for the Seyfert 2 luminosity function. The luminosity function of Seyfert 1s covers brighter Hα luminosities than that of Seyfert 2s. At low luminosity, the relative ratio of Seyfert 1s to Seyfert 2s is nearly 1. However the luminosity function of Seyfert 2s drops off quickly as luminosity increases. As a result, at high luminosities, the relative ratio of Seyfert 1s to Seyfert 2s becomes much larger than 1. This may be related to the fact that there are very few type II quasars observed while low-luminosity Seyfert 2s are as common as their broad-line counterpart. The change of the relative ratios happen at approximately Hα luminosity $10^7 L_\odot$. This result is still quite fresh and there is a lot of work to do to fully understand the two luminosity functions. For further detailed results, please refer to Hao et al. (2004).

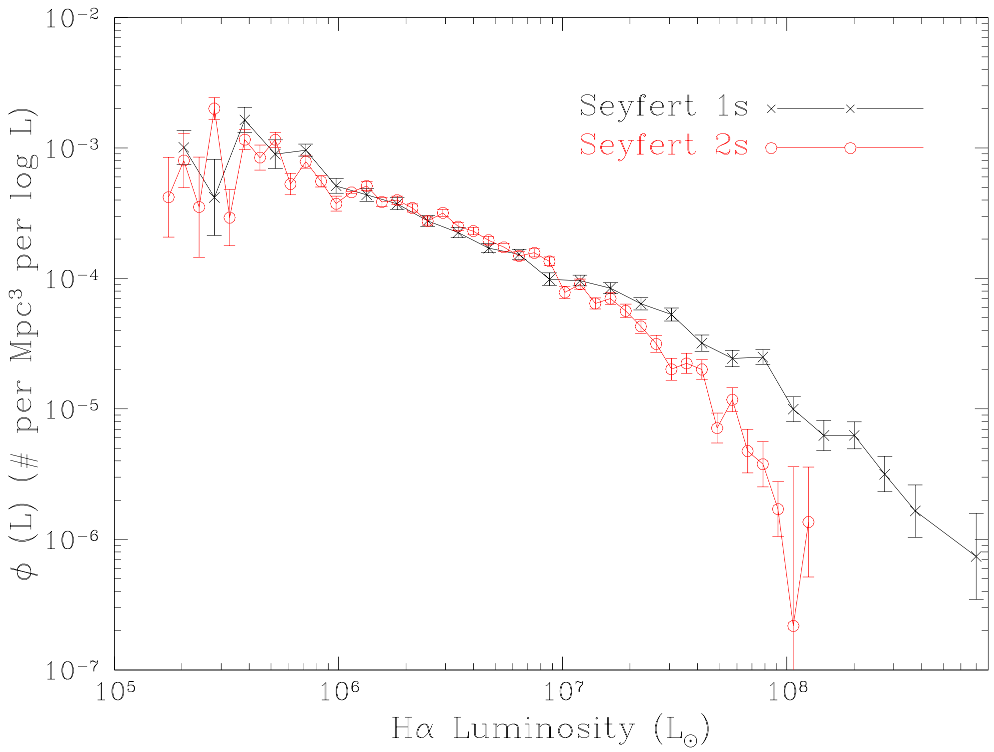

Fig. 1.6. The luminosity function for Seyfert 1s (black, crosses) and Seyfert 2s (red, open circles).

1.5.4 Luminosity Function Justification

To check if the derived luminosity function is correct, we apply the following test. Since the idea and procedure for Seyfert 1s and Seyfert 2s are the same, I will focus here only on Seyfert 2s. We divide the Seyfert 2s included in the luminosity function calculation into several subgroups by their redshifts. For each group, we evaluate the AGN luminosity distribution expected from the luminosity function measured, and it can be described as:

$$F_{expect}(L) = \sum_i \frac{p_i(L)\Phi(L)}{\int_0^\infty \Phi(L')p_i(L')dL'}$$

Where the summation includes all AGNs within each group. If the luminosity function is correct, the real luminosity distribution *in each group* should be similar to the expected luminosity function. Figure 1.7(a) demonstrates this test for the Seyfert 2 luminosity function. It can be seen that the observed and expected luminosity distribution agree well with each other, which implies that our luminosity function result is correct. The same test is also done for Seyfert 1s and similar result is got.

A similar test is done by dividing AGNs not by their redshifts but by their host galaxy luminosities. This test will directly tell us if our assumption that the nuclear luminosity is independent of the host galaxy luminosity is reasonable. Since the expected luminosity distribution is evaluated from the luminosity derived based on this assumption, any relationship

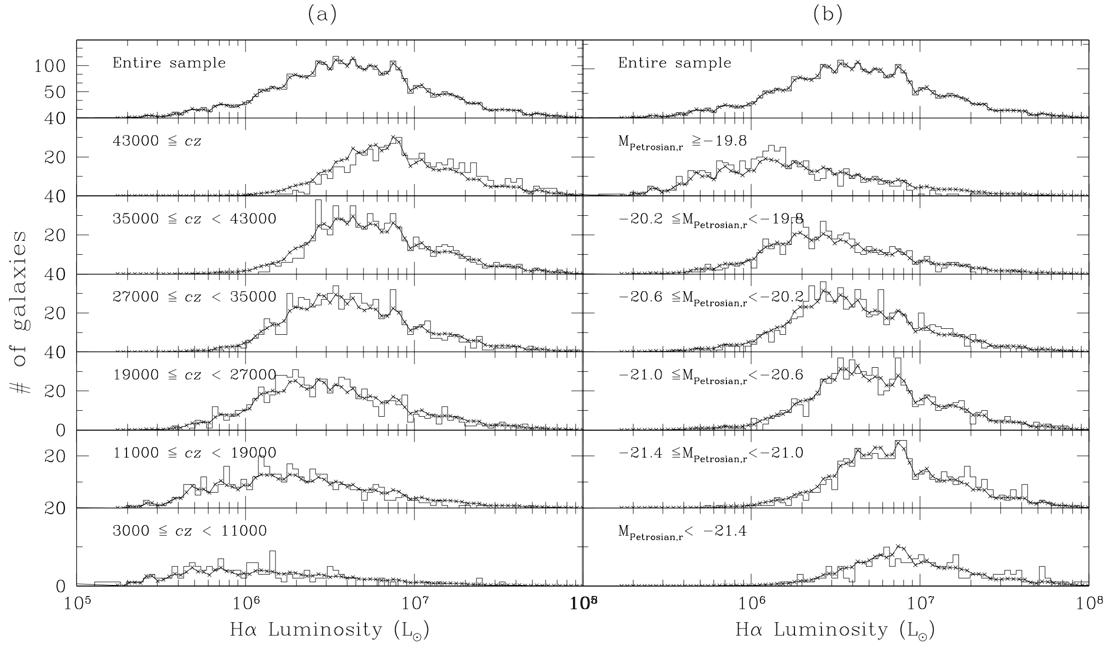

Fig. 1.7. The real luminosity distribution for Seyfert 2s (histogram) compared with the expected luminosity distribution (solid line) calculated from the Seyfert 2 luminosity function (Figure 1.6) for different subgroups of the Seyfert 2s sample. *Left*(a): the entire Seyfert 2s sample (shown in the top panel) is grouped by their redshifts. *Right*(b): the entire Seyfert 2s sample is grouped by their absolute *r* band Petrosian magnitudes.

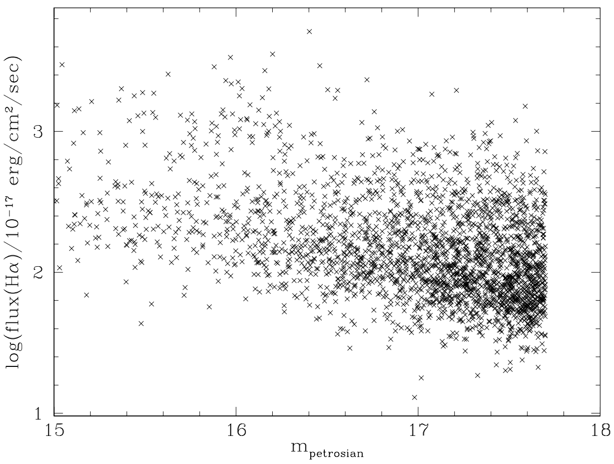

Fig. 1.8. The r-band Petrosian magnitude vs. the Hα emission-line flux for the Seyfert 2s sample.

between the two luminosities would cause systematic differences between the expected and the real luminosity distributions in different host galaxy luminosity subgroups. Figure 1.7(b) gives the result of this test for Seyfert 2s. As can be seen, the two distributions match up in every subgroup, which confirms that indeed the nuclear luminosity is independent of the host galaxy luminosity. This is also reflected in Figure 1.8, in which we plot Hα emission-line flux against the host galaxy Petrosian magnitude for Seyfert 2s. There is an apparent selection effect: few objects exist with bright host galaxy but weak nuclei (lower-left part of the plot). This is because the nuclear features in this kind of galaxies are overwhelmed by their host galaxies, therefore they are undetected. Apart from this, no visible relationship between the two variables is seen. The independence of the two variables implies that the principal quantity that determines AGN luminosity is the Eddington ratio, not the black hole mass. A similar result is also found for Seyfert 1s sample.

1.6 Conclusions

We have applied a thorough spectroscopic identification of AGNs from the Sloan Digital Sky Survey. From 83,839 SDSS galaxy spectra, we have identified 1279 broad and 2694 narrow-line AGNs, making up by far the largest low-redshift AGN sample selected by their emission line features. To accurately measure the emission line parameters, we have applied careful stellar subtraction to the original galaxy spectra. The stellar templates used in the subtraction are mainly constructed from Principal Component Analysis of pure absorption-line galaxies. The distribution of the Hα FWHM for the emission line galaxies in SDSS is bimodal, with two populations separating at FWHM(Hα)$\sim$ 1,200 km/s. This gives a natural selection criterion for broad-line AGNs. Narrow-line AGNs are identified by their locations in the traditional emission line ratio diagnostic diagrams. We measure the luminosity functions for these AGNs with the detection probability function evaluated through a Monte-Carlo simulation. The comparison of the luminosity function of Seyfert 1s and Seyfert 2s indicates that the relative ratio of Seyfert 1s to Seyfert 2s changes with luminosity, from nearly 1 at low Hα luminosity to a relatively large number at high Hα luminosity. Finally, we demonstrated that the nuclei luminosity is not strongly correlated with the host galaxy luminosity, which means that the principal quantity that determines AGN luminosity is the Eddington ratio, not the black hole mass. This result provides important clues and constraints on AGN formation and evolution theories.

Acknowledgments. Funding for the creation and distribution of the SDSS Archive has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, Princeton University, the United States Naval Observatory, and the University of Washington.

References

Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Blanton, M. 2000, ApJ, 544, 63

Efstathiou, G., Ellis, R. S., & Peterson, B. A. 1988, MNRAS, 232, 431
Gunn, J. E., et al. 1998, AJ, 116, 3040
Hao, L., et al. 2004, in preparation
Ho, L. C., Filippenko, A. V., & Sargent, W. L. W. 1997, ApJS, 112, 315
Kauffmann, G., et al. 2004, MNRAS, in press (astro-ph/0205070)
Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, ApJ, 556, 121
Koranyi, D. M., & Strauss, M. A. 1997, ApJ, 477, 36
Saraiva, M. F., Bica, E., Pastoriza, M. G., & Bonatto, C. 2001, A&A, 376, 43
Strauss, M. A., et al. 2002, AJ, 124, 1810
Veilleux, S., & Osterbrock, D. E. 1987, ApJS, 63, 295
York, D. G., et al. 2000, AJ, 120, 1579