Chin. J. Astron. Astrophys. Vol. 6 (2006), No. 2, 197–209
(*http://www.chjaa.org*)

Chinese Journal of
Astronomy and
Astrophysics

A Catalog of Luminous Infrared Galaxies in the *IRAS* Survey and Second Data Release of SDSS *

Chen Cao[1,2], Hong Wu[1], Jian-Ling Wang[1,2], Cai-Na Hao[1,2], Zu-Gan Deng[3,1], Xiao-Yang Xia[4] and Zhen-Long Zou[1]

[1] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012; *caochen@bao.ac.cn*
[2] Graduate School of Chinese Academy of Sciences, Beijing 100049
[3] College of Physical Sciences, Graduate School of Chinese Academy of Sciences, Beijing 100049
[4] Department of Physics, Tianjin Normal University, Tianjin 300074

Abstract We selected a sample of luminous infrared galaxies by cross-identification of the Faint Source Catalogue (FSC) and Point Source Catalogue (PSC) of the *IRAS* Survey with the Second Data Release of the SDSS. The size of our sample is 1267 for FSC and 427 for PSC by using the 2σ significance level cross-section. The "likelihood ratio" method is used to estimate the individual's reliability and for defining two more reliable subsamples (908 for FSC and 356 for PSC). A catalog of infrared, optical and radio data is compiled and will be used in further work. Some statistical results show that luminous infrared galaxies are quite different from ultra-luminous infrared galaxies. The AGN fractions of galaxies at different infrared luminosities and the radio–infrared correlations are consistent with the previous studies.

Key words: catalogs — galaxies: statistics — infrared: galaxies

1 INTRODUCTION

The research of Luminous Infrared Galaxies (LIGs, the galaxies with infrared luminosity ($L_{\rm IR}$, 8–1000 μm) higher than $10^{11}L_{\odot}$) began after the success of the first mid- to far-infrared all-sky survey carried out in 1983 by the Infra-Red Astronomical Satellite (*IRAS*). The physical properties of the LIGs, especially the Ultra-Luminous Infrared Galaxies (ULIGs, $L_{\rm IR} > 10^{12}L_{\odot}$) were studied by using the *IRAS* infrared data and the follow-up optical observations (POSS, DSS, *HST*, VLT ...), such as the analyses of the Bright Galaxy Sample (BGS, Soifer et al. 1987b), the optical spectroscopy of LIGs (Kim et al. 1995; Veilleux et al. 1995), the statistical study of the spectra of very luminous IRAS galaxies (Wu et al. 1998a, b), the IRAS 1 Jy Survey of ULIGs (Kim et al. 1998a, b) and the Point Source Catalog redshift survey (PSCz, Saunders et al. 2000). From the previous studies it has been found that most of the ULIGs are in interaction/merger systems (Zou et al. 1991; Sanders et al. 1988; Kim et al. 1995; Lawrence et al. 1989) and with a high AGN fraction (Kim et al. 1995, 2002; Wu et al. 1998a, b). There is a possible evolution path (Sanders et al. 1988; Sanders & Mirabel 1996) from galaxy mergers to quasi-stellar objects (QSOs) and elliptical galaxies, which supports the hierarchical galaxy formation theory (Cole et al. 2000).

*Supported by the National Natural Science Foundation of China.

LIGs with $L_{\rm IR} \sim 10^{11}$ – 10^{12} $L_{\odot}$ are quite different from the ULIGs in their morphologies and spectral features. Recent studies of distant LIGs ($0.4 < z < 1.2$, Zheng et al. 2004) showed that there are many massive disks which have been forming a large fraction of their stellar mass since $z = 1$, and most of their central parts were formed prior to the formation of their disks. Despite their importance there has not been a large and reliable sample of LIGs for statistical analyses, so many of their physical properties are still unclear. The role of LIGs and ULIGs in the formation and evolution of the galaxies is still an open problem.

In order to study the properties of the LIGs in more detail, we need a large sample which has both infrared and optical data for analysis. The Sloan Digital Sky Survey (SDSS) was chosen for a cross-identification with *IRAS* data because of its large sky coverage (~ 2627 deg^2 for spectroscopic targets of the second data release) and high spectral signal-to-noise (S/N) ratio and spectral resolution ($R \sim 1800$). Some authors have studied the optical properties of *IRAS* galaxies using the SDSS data (Goto 2005b; Pasquali et al. 2005), but the cross-identification between optical and infrared catalogs resulted previously is relatively simple (using only a fixed circle) for a reliable sample selection and they did not present a complete catalog for further analyses.

The structure of the present paper is as follows: in Section 2 we give a simple description of the data and the cross-identification between *IRAS* and SDSS; in Section 3 we use the "likelihood ratio" method for detailed identifications for our sample and estimate its reliability; in Section 4 we describe our catalog; we carry out a statistical census based on a selected subsample in Section 5. Finally a summary is given in Section 6. We adopt cosmological parameters $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\rm m} = 0.3$, and $\Omega_{\Lambda} = 0.7$, throughout this paper.

2 DATA DESCRIPTION AND SAMPLE SELECTION

2.1 *IRAS* Faint Source Catalog and Point Source Catalog

The Infra-Red Astronomical Satellite (*IRAS*) was launched in 1983 (Neugebauer et al. 1984; Soifer et al. 1987a) and scanned almost the whole sky in mid- and far-infrared (12, 25, 60 and 100 μm) wavebands. The Faint Source Catalog (FSC, |b| > 10, Version 2.0, Moshir+ 1989) was released after the Point Source Catalog (PSC, Version 2.0, IPAC 1986). It contains data for 173044 point sources in unconfused regions with flux densities typically above 0.2 Jy at 12, 25 and 60 μm, and above 1.0 Jy at 100 μm, reaching roughly one-magnitude deeper in sensitivity than the PSC. The catalogues (both the FSC and PSC) give the *IRAS* sources' four band flux densities and qualities, the positions of the sources, and other useful parameters. Sources in the catalogues all have large positional uncertainties which are described by an "error ellipse", that specifies the uncertainties along (in-scan) and cross (cross-scan) the *IRAS* scan direction, and the position angle of the major axis. The FSC is deeper than the PSC but may be contaminated by foreground and background sources, the PSC is shallower but can be used for a comparison with previous results (e.g., the PSCz). Therefore, we use them separately to make up our sample and carry out separate statistical analyses of them.

2.2 SDSS-DR2 Data

The Sloan Digital Sky Survey (SDSS, York et al. 2000) contains an imaging survey of the northern sky in the five bands, u, g, r, i and z and a spectroscopic target survey performed by multi fibers. The Second Data Release (DR2, Abazajian et al. 2004, Version v2_20040928_1505) was released in 2004. The SDSS-DR2 spectroscopic target survey covers about 2627 deg^2 of the sky, including about 260490 galaxies, 32241 quasars, 3791 high-z ($z > 2.3$) quasars and others objects. For the study of the detailed spectral properties of LIGs (such as their emission lines), we selected only those SDSS-DR2 spectroscopic targets with redshifts greater than 0.001 (to reject stars) and high redshift confidence (zConf > 0.9) in the cross-identification. Finally we obtained 268202 sources from the SDSS datasets as our candidates for the cross-identification with the *IRAS* catalogues.

2.3 Cross-identification between the *IRAS* and SDSS

We use the *IRAS* (FSC and PSC, separately) error ellipse as the cross-section (the SDSS's position uncertainties are negligible in comparison) to do the cross-identification with the SDSS spectral positions. Two RMS uncertainty (2σ) significance level was chosen for a high level confidence and more complete sample selection. The SDSS spectral redshift and *IRAS* flux densities were then used to calculate the infrared luminosity ($L_{\rm IR}$) of the matched sources. Due to the fact that the $12\,\mu$m and $25\,\mu$m flux densities of the objects are mostly "upper limits" (flux quality = 1), we calculate the far-infrared luminosity (Helou et al. 1988; Sanders & Mirabel 1996) and then convert it to the total infrared luminosity (1–1000 μm, Calzetti et al. 2000)[1]

$$F_{\rm FIR} = 1.26 \times 10^{-14}\{2.58 f_{60} + f_{100}\}[\mathrm{W\ m^{-2}}]\,, \tag{1}$$

$$L_{\rm FIR} = 4\pi D_{\rm L}^2 F_{\rm FIR}[L_\odot]\,, \tag{2}$$

$$L_{\rm IR}(1-1000\,\mu\mathrm{m}) = 1.75 L_{\rm FIR}\,, \tag{3}$$

where f_{60}, f_{100} are the *IRAS* flux densities in Jy at 60 and 100 μm, respectively. Then the LIGs ($L_{\rm IR} \geq 10^{11} L_\odot$) were chosen as our sample objects, and the number of sources is 1267 for FSC and 427 for PSC.[2] From this sample we compiled a catalog (which will be described in Section 4) and performed detailed identifications and further analyses. Figure 1 shows the sky coverage of our sample (both FSC and PSC) in equatorial coordinates, representing almost all the SDSS-DR2 spectroscopic survey regions.

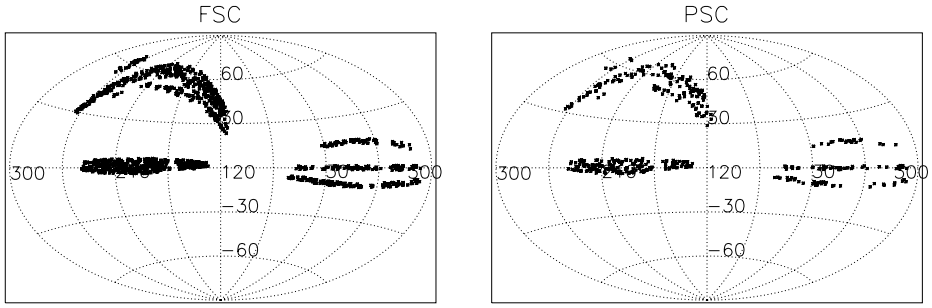

Fig. 1 Distribution on the sky of the objects in our sample. This is an Aitoff projection in equatorial coordinates. Left: FSC sample; Right: PSC sample.

2.4 VLA-FIRST Data

The NRAO Very Large Array (VLA) Faint Images of the Radio Sky at Twenty-centimeters (FIRST) data (Becker et al. 1995) are used here for studying the radio properties of our sample. The FIRST survey was a project designed to produce a radio equivalent of the Palomar Observatory Sky Survey (POSS) over 10^4 deg^2 of the North and South Galactic Caps. The FIRST Survey Catalog (White et al. 1997, from the 1993 through 2002, contains $\sim$ 811 000 sources and covers $\sim$ 9030 deg^2) including peak and integrated flux densities and the size information was generated from the coadded images. The individual sources have 90% confidence error circles of

[1] The contribution to the total infrared luminosity from the 1–8 μm regime is expected to be of the order of a few percent (Calzetti et al. 2000).

[2] Note that the objects with 60 μm flux quality =1 have been rejected. We did not treat the objects with 100 μm upper limits because it does not affect much on the calculation of $L_{\rm IR}$ (see Section 5.2 for details).

radius $< 0.5''$ at the 3 mJy level and $< 1''$ at the survey threshold ($\sim$ 1 mJy). The survey area was chosen to coincide with that of the SDSS First Data Release (DR1) and $\sim 50\%$ of the optical counterparts to FIRST sources will be detected. We use the FIRST Survey Catalog updated at 2003 April 11 to perform the cross-identification with the objects in our sample.

We match our sample's SDSS spectral positions with the VLA FIRST positions using a $2''$ searching radius and find that there are 624 objects in FSC and 258 in PSC that are contained in the FIRST catalog. This result means that the radio flux densities of these sources are all above the FIRST's threshold (about 1 mJy). Thus they have a higher probability to be true IR sources because of the (far-) infrared to radio correlation (Helou et al. 1985, 1993; Condon 1992; Ivezić 2002, and will be discussed in Section 5).

2.5 Reliability and Completeness

Due to our large 2σ cross-sections for the cross-identification, there are also some SDSS objects which are not really the IR sources being selected as our sample objects because of the contamination of foreground and/or background sources. So we calculate the random probability that SDSS-DR2 spectroscopic targets fall into the *IRAS* 2σ error ellipse by assuming that the SDSS targets are uniformly distributed over the 2627 deg^2 sky and the mean *IRAS* 2σ error ellipse area is about 0.56 arcmin^2 for the LIGs. The random probability is about 4.32% for FSC sample and 5.02% for PSC and hence our whole sample's reliability is about 95.68% (FSC) and 94.98% (PSC) ($R = 1 - N_{\rm random}/N_{\rm real}$).

The completeness of our sample can be estimated from the 2σ error ellipse cross-section, and the incompleteness introduced by this term alone is about 10% assuming a Gaussian distribution. In addition, it also may be affected by several other factors:

(1) We only selected SDSS targets with high confidence redshift (zConf > 0.9) as our candidates, which will lead to some targets without high quality redshift estimates being rejected. The incompleteness increases from 1% to 6%, from the bright objects to the faint ones.
(2) Because of the target magnitude limit of the SDSS spectroscopic survey (Petrosian mag_r $\leq$ 17.77 for main galaxies and PSF mag_i $\leq$ 19.1 for quasars), there are also some optically faint LIGs which could not be included in the SDSS spectroscopic survey. So they are missed mainly due to their relatively higher redshift or serious obscuration by dust.
(3) There are also missed galaxies due to the lack of fibers in dense regions, spectroscopic failures, and fiber collisions, which can be defined by the sampling rate: $\tilde{\rm f}_{\rm t} \sim 0.92$ on average (Blanton et al. 2001).

3 "LIKELIHOOD RATIO" METHOD

It is rather difficult to determine whether the matched SDSS targets are really the infrared objects or not. So we use the "Likelihood Ratio" (LR) method (Sutherland & Saunders 1992) to calculate the probability of the "true" cross-identification for each matched SDSS object.

The LR method is defined as that the cross-identification probability between two observed sources is (assume that the errors are Gaussian in common)[3]

$$LR = \frac{Q(\leq m_{\rm i}) \exp(-r^2/2)}{2\pi\sigma_{\rm a}\sigma_{\rm b} n(\leq m_{\rm i})} \,. \tag{4}$$

In Equation (4) r is the "normalized distance",

$$r^2 = \frac{(a1-a2)^2}{\sigma_{a1}^2+\sigma_{a2}^2} + \frac{(b1-b2)^2}{\sigma_{b1}^2+\sigma_{b2}^2} \,, \tag{5}$$

[3] Note that the cross-scan errors for faint galaxies of *IRAS* are less Gaussian (*IRAS* Explanatory Supplement VII, Analysis of Processing C. Positional Accuracy), but this does not affect much on our statistical results. So we use the Gaussian assumption and the LR method here and will try to improve it in further work.

where ($a1$, $b1$) and ($a2$, $b2$) are the positions of the sources, σ is the standard deviation and $n(\leq m_{\rm i})$ the local surface density of objects (galaxies) brighter than the candidate. $Q(\leq m_{\rm i})$ is the multiplicative factor in the numerator which represents the a priori probability that a "true" optical counterpart brighter than the flux limit exists amongst the identifications, and for simplicity we set $Q = 1$ in this work.

For our sample, SDSS position uncertainties can be neglected compared with the *IRAS*'s large error ellipse. In this work, we refer to the *IRAS* uncertainty ellipse major axis (UncMaj) as $\sigma_{\rm a}$, minor axis (UncMin) as $\sigma_{\rm b}$ and the position of the SDSS object in the *IRAS* 2σ error ellipse (in units of σ, from 0 to 2) as r. We use the SDSS photometric targets to calculate $n(\leq m_{\rm i})$

$$n(\leq m_{\rm i}) = \frac{N(\leq m_{\rm i})}{4\pi\sigma_{\rm a}\sigma_{\rm b}}\,, \tag{6}$$

where $N(\leq m_i)$ stands for the number of galaxies with r band magnitude less than or equal to the candidate's in the corresponding *IRAS* 2σ error ellipse. Then we can obtain the LR formula for our sample

$$\mathrm{LR} = \frac{2\exp(-r^2/2)}{N(\leq m_{\rm i})}\,. \tag{7}$$

We calculate all of our samples' likelihood ratio values by using the SDSS photometric data (r band Petrosian magnitude for galaxies and i band PSF magnitude for QSOs). Then a random sample is selected for estimating the reliability of each object (use the method developed by Lonsdale et al. 1998; Rutledge et al. 2000; Masci et al. 2001), which is used to assess the cross-identification probability and select a more reliable subsample. We also calculate the LRs and reliabilities for the PSCz sample (Saunders et al. 2000, all these optical targets selected from the PSC are identified as "true" IR objects) overlapped with our PSC sample for a comparison. The reliability distributions of the FSC, PSC and PSCz sample are shown in Figure 2.

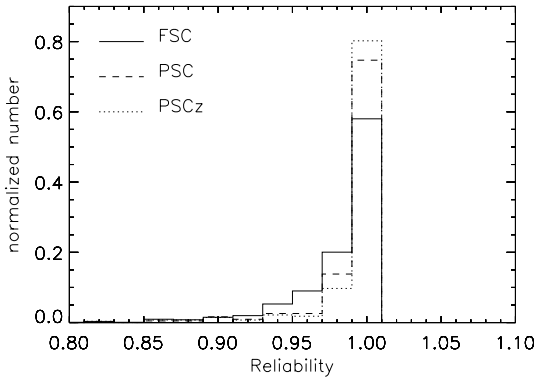

Fig. 2 Distributions of our sample's reliabilities. The solid line shows the FSC sample, dashed line the PSC sample, and dotted line the PSCz sample.

4 THE CATALOG

We present a catalog (in ascii) for our sample of LIGs, containing the information given in the *IRAS*, SDSS-DR2 and FIRST. The parameters of the catalog are as follows: [4]

[4] The Catalog will be put on the web, use this URL:http://www.chjaa.org/2006_6_2_p197/catalog.tar.gz.

The *IRAS* data (f(p)sciras.cat): the *IRAS* (FSC and PSC) name; *IRAS* RA and DEC; the error ellipse major (UncMaj), minor axis (UncMin)[5] and position angle; 12, 25, 60 and 100 μm flux densities and qualities; and the calculated infrared luminosity using the SDSS spectral redshift.

The SDSS-DR2 photometric data (f(p)scsdssphoto1(2).cat): the SDSS ObjID; Photometric RA and DEC; objType and probPSF; SDSS five bands modelMag, psfMag, fiberMag, petroMag and their errors; Galactic extinctions; petroR50 and petroR90 for band r.

The SDSS-DR2 spectroscopic data (f(p)scsdssspec1.cat): the SDSS SpecObjID; Spectroscopic RA and DEC; spectral redshift and its error; eclass and eCoeff; zWarning and zStatus; SpecClass, mjd, plate, and fiberID.

The SDSS-DR2 emission line data (f(p) scsdssspec2.cat, from MPA-SDSS: www.mpa-garching.mpg.de/SDSS, Version 5.0_4, Tremonti et al. 2004): the Hα, Hβ, [OII]$\lambda\lambda$3727, 3729, [OIII]λ5007, [NII]λ6584, [SII]$\lambda\lambda$6716, 6731 and [OI]λ6300 emission line fluxes and flux errors; the corresponding Equivalent Widths (EQWs) and errors. Based on these data, we classify our sample into several spectral types:

a) Galaxies without apparent emission lines (NoE for short) chosen by the criterion: Hα EQW > -5 Å.[6];
b) QSOs/Seyfert 1s (S1) being those with Broad Line Regions (BLRs) and also classified as QSOs by the SDSS pipeline (specClass = 3);
c) The classification of narrow emission line galaxies (Seyfert 2s, LINERs and HII regions) using the emission line fluxes ratios, [OIII]λ5007/Hβ, [NII]λ6584/Hα, [SII](λ6716+λ6731)/Hα, [OI]λ6300/Hα (Osterbrock 1985,1989; Wu et al. 1998b; Kauffmann et al. 2003c; Kewley et al. 2001). Specifically, for Seyfert 2s (S2): [OIII]/H$\beta \geq 3$; For LINERs (L): [NII]/H$\alpha > 0.6$, [SII]/H$\alpha > 0.4$, [OI]/H$\alpha > 0.05$ and [OIII]/H$\beta < 3$; For HII galaxies (H): [NII]/H$\alpha < 0.6$, [SII]/H$\alpha < 0.4$, [OI]/H$\alpha < 0.05$ and [OIII]/H$\beta < 3$; The mixed types (LH: Mixture of LINERs and HIIs) are those located at the border of different spectral populations. The mixed type galaxies could be a transitional phase from HII galaxies to AGNs (Wu et al. 1998b). Also, there are some galaxies which are not in the MPA's emission line catalog, so we classify them as Unknown (?). We will discuss this category in detail in Section 5.3.

The VLA FIRST radio data (f(p)scfirst.cat): The VLA FIRST data (described in Section 2.4) contains the FIRST name; FIRST RA and DEC; peak and integrated flux densities at 1.4 GHz; the local noise estimate; major and minor axis (FWHM), position angle; fitted MajAxis, MinAxis and PA before deconvolution; name of the coadded image containing the source; and based on the cross-identification we give a "flags" for our sample: 0 stands for the case that the SDSS object is correlated with a FIRST source within $2''$ and 1 stands for the contrary case.

We give each source a new serial number separately for the FSC and PSC samples, which will be used in our analysis.

The main catalog (f(p)sc_main.cat) contains only the most important information we need, which includes the source number, the likelihood ratio (LR) and the Reliability we calculated in Section 3, the *IRAS* name, the infrared luminosity, redshift, SpecObjID, Spectroscopic RA and DEC, SpecClass, ObjID, modelMag_r, extinction_r, petroMag_r, the FIRST flag, the SDSS object's position in the *IRAS* error ellipse (in units of σ), the spectral types and the sign of the same sources across the two sample (FSC and PSC).

5 ANALYSES AND RESULTS

5.1 Subsample Selection

For the purpose of high confidence analyses we need a subsample with relatively high reliabilities. From a comparison between our sample and the random sample (discussed in Section 3 and shown in Fig. 2), we get a selection criterion, "Reliability ≥ 0.98" for a relatively high cross-identification

[5] Note that the UncMaj and UncMin in the PSC stand for 1.96σ significance level.
[6] Absorption lines have a positive sign.

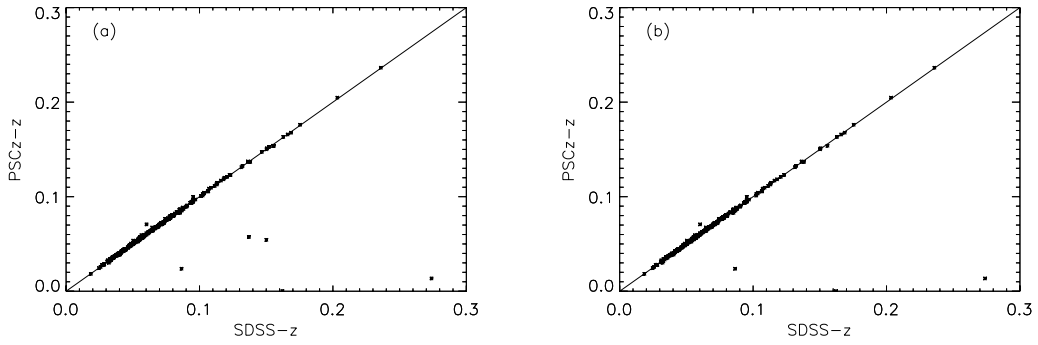

Fig. 3 A comparison between the redshifts of (a) PSC whole sample and PSCz; (b) PSC subsample and PSCz.

probability. With this criterion we obtain an FSC subsample of 908 objects and a PSC subsample of 356. From a comparison of the redshifts in our PSC sample, PSC subsample and the PSCz sample of the same *IRAS* source (Fig. 3), we find that our subsample (at least the PSC) is more reliable because there are only two wild points in the scatter diagram. We also estimated our subsample's completeness from the LR distribution of the PSCz sample and found it to be about 86.69% if we used the same selection criterion.

5.2 Basic Statistical Properties

The redshift distributions (and partial distributions for the LIGs and ULIGs) of our subsamples are shown in Figures 4 and 5. The number of LIGs ($N_{\rm LIGs}$, with $L_{\rm IR} \sim 10^{11}$ – $10^{12} L_{\odot}$) is 873 for the FSC and 334 for the PSC, and $z_{\rm median} \sim 0.08$ (FSC) and 0.05 (PSC). For the ULIGs (with $L_{\rm IR} > 10^{12}\ L_{\odot}$), $N_{\rm ULIGs}$ is 35 (FSC) and 22 (PSC), and $z_{\rm median} \sim 0.18$ (FSC) and 0.17 (PSC), ~ 0.1 higher than the LIGs. The ratio $N_{\rm ULIGs}$: $N_{\rm LIGs}$ is 0.04 for the FSC and a higher value 0.07 for the PSC. For a comparison of the infrared luminosities derived from the FSC and the PSC (see Fig. 6), we find that the $L_{\rm IR}$ derived from the FSC is consistent with that from the PSC according to the formula given in Section 2.3.

The color (u-r) distributions of our subsamples are shown in Figure 7. Compared with the color separation of galaxy types described by Strateva et al. (2001), our result shows higher u-r values. Serious dust extinction in the LIGs, and even more serious extinction in the ULIGs, may be responsible for the redder color in our subsamples.

5.3 AGN Fraction

Throughout this paper, the term AGNs cover Seyfert 1s, Seyfert 2s, LINERs, and the Mixed types (S1+S2+L+LH, the spectral types are described in Section 4). The BPT (Baldwin et al. 1981) diagrams for classifying the narrow emission line galaxies (Seyfert 2s (S2), LINERs (L), HIIs (H) and Mixed types) are shown in Figure 8. The number and fraction of each type are listed in Tables 1 and 2. Figure 9 displays the composition of the various types (S1, S2 etc., excluding the "unknowns") in three $L_{\rm IR}$ intervals, separately for our two subsamples. Note that we have made a volume correction by giving each object a weight equal to the inverse of its maximum visibility volume: $1/V_{\rm max}$ (Schmidt 1968; Kauffmann et al. 2003ab), with a magnitude and flux cutoff for correcting the selection biases. We calculate $V_{\rm max}$ by

$$\log D_{\rm l}({\rm max})_{\rm SDSS} = \frac{{\rm mag}_{\rm lim} - {\rm mag}}{5} + \log D_{\rm l}(z), \tag{8}$$

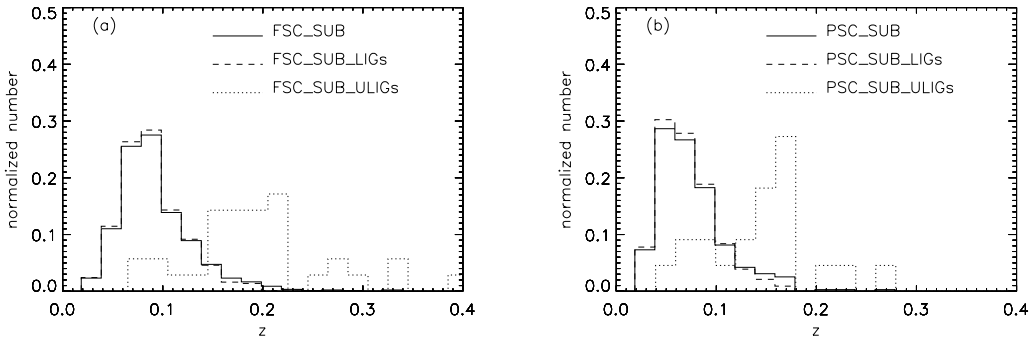

Fig. 4 Redshift distribution of our LIGs subsample. (a): FSC subsample; (b): PSC subsample. The solid lines are for the whole subsample, dashed lines for LIGs ($L_{\rm IR} \sim 10^{11}$ – $10^{12} L_{\odot}$), and dotted lines for ULIGs ($L_{\rm IR} > 10^{12} L_{\odot}$).

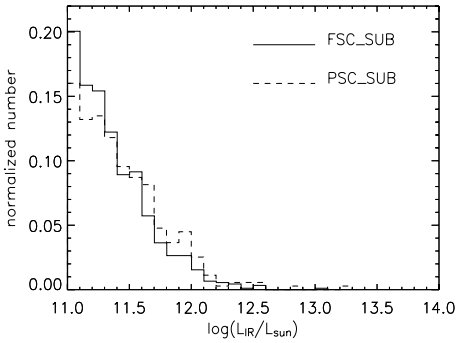

Fig. 5 Infrared luminosity distribution of LIGs in our FSC subsample (solid line) and in our PSC subsample (dashed line).

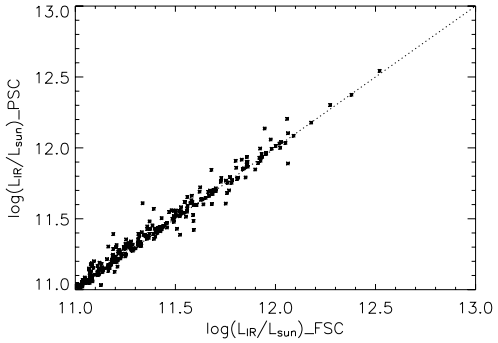

Fig. 6 A comparison between the infrared luminosities derived from the FSC and the PSC subsamples, all with 60 μm flux qualities = 2 or 3.

$$D_{\rm l}({\rm max})_{\rm IRAS} = D_{\rm l}(z)\Big(\frac{f60}{f60_{\rm lim}}\Big)^{1/2}. \quad (9)$$

In this equation $\rm mag_{lim}$ is the SDSS magnitude cutoff (Petrosian mag_r = 17.5), and $f60_{\rm lim}$ the *IRAS* 60 μm flux cutoff (0.3 Jy for FSC, 0.6 Jy for PSC). Then $D_{\rm l}({\rm max})$ for our estimation is the minimum of $D_{\rm l}({\rm max})_{\rm SDSS}$ and $D_{\rm l}({\rm max})_{\rm IRAS}$, so: $V_{\rm max} = 4/3\pi(D_{\rm l}^3({\rm max})/(1+z)^3)$.

The AGN fraction in our subsamples increases with the infrared luminosity, from $\sim 45\%$ to 80% as $L_{\rm IR}$ increases from 10^{11} to $10^{13} L_{\odot}$. This is in agreement with the previous results that the AGN fraction increases from the LIGs to ULIGs, from 47% to 70%–75% (Kim et al. 1995; Veilleux et al. 1995, 1999) and 56% to 82% (Wu et al. 1998b). From Tables 1 and 2 we also find that some galaxies without apparent emission lines (NoE) have a high $L_{\rm IR}$, especially for the PSC subsample (due to their relatively higher $L_{\rm IR}$). These galaxies may either: a) have low S/N ratios or bad spectra; or b) be one member of a galaxy pair or group, and the large amount of infrared emissions may come from its companion; c) have a late stage merger feature and e(a) spectral feature (Poggianti & Wu 2000) or E+A feature, which indicates a post-starburst phase (Zabludoff et al. 1996; Yang et al. 2004; Goto 2005a).

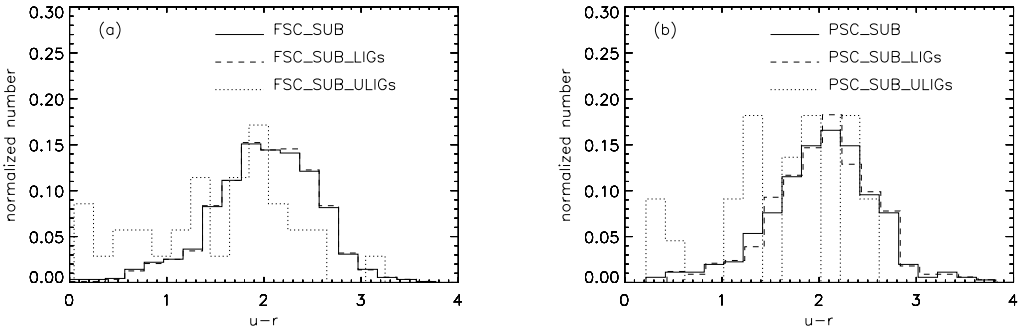

Fig. 7 Color (u-r) distributions of LIGs (solid lines) and ULIGs (dotted lines) in (a) the FSC subsample, (b) the PSC subsample.

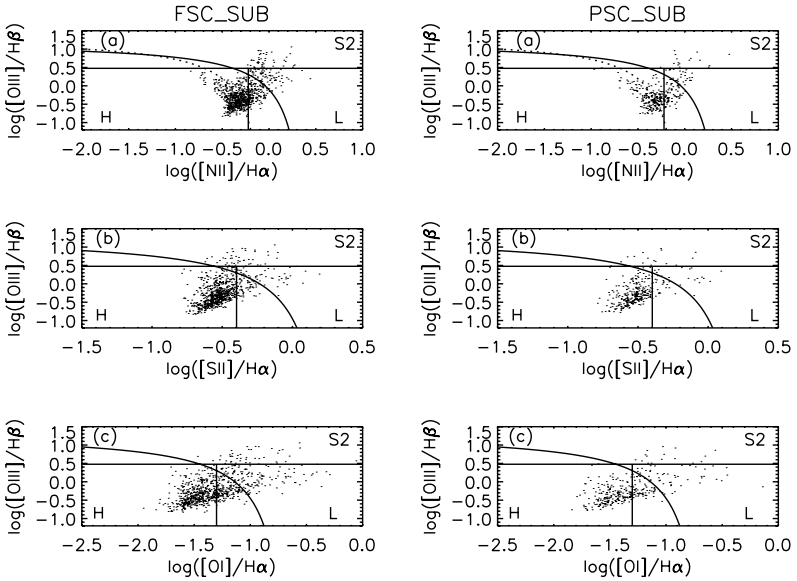

Fig. 8 BPT diagrams of LIGs in our two subsamples. The straight lines are the criterion we use in this paper to separate Seyfert 2s (S2), LINERs (L) and HIIs (H). The solid curve is the criterion given by Kewley et al. (2001) for separating starbursts and AGNs, and the dashed curve in (a) is the criterion given by Kauffmann et al. (2003c).

5.4 Infrared to Radio Correlation

The infrared–radio correlation in our two subsamples is shown in Figure 10. We calculate $L_{60\,\mu\mathrm{m}}$ and $L_{1.4\,\mathrm{GHz}}$ using formulas (Yun et al. 2001)

$$\log L_{60\,\mu\mathrm{m}}(L_\odot) = 6.014 + 2\log D + \log S_{60\,\mu\mathrm{m}}, \tag{10}$$

and

$$\log L_{1.4\,\mathrm{GHz}}(\mathrm{W\ Hz^{-1}}) = 20.08 + 2\log D + \log S_{1.4\,\mathrm{GHz}}\,, \tag{11}$$

Table 1 Spectral type distribution in three infrared luminosity intervals in our FSC subsample. Error for the AGN fraction based on Poisson statistics.

Spectral Type	$\log L_{\rm IR}\ (L_\odot) \sim$		
	11.0–11.5	11.5–12.0	> 12.0
S1[a]	0.94%(6)[b]	0.75%(2)	7.87%(4)
S2	5.22%(22)	3.89%(5)	0.00%(0)
L	9.05%(27)	8.41%(10)	5.53%(1)
LH	30.49%(100)	36.52%(44)	64.71%(5)
H	49.98%(183)	47.70%(58)	21.88%(1)
NoE	4.33%(13)	2.73%(5)	0.00%(0)
Total	351	124	11
AGN	45.69±3.67%(155)	49.57±6.35%(61)	78.12±24.70%(10)

[a] The spectral types: S1, S2, L, LH, H and NoE stand for the Seyfert 1s, Seyfert 2s, LINERs, Mixed types and HIIs, respectively, as described in Section 4.
[b] The volume corrected fraction of different spectral types in each $L_{\rm IR}$ bin. The number of each type galaxies is in the bracket.

Table 2 Spectral type distribution in three infrared luminosity intervals in our PSC subsample

Spectral Type	$\log L_{\rm IR}(L_\odot) \sim$		
	11.0–11.5	11.5–12.0	> 12.0
S1	1.59%(3)	2.59%(2)	3.69%(3)
S2	6.31%(7)	6.43%(3)	0.00%(0)
L	7.34%(9)	4.74%(2)	0.00%(0)
LH	37.53%(47)	33.83%(22)	74.94%(6)
H	45.53%(70)	50.06%(30)	21.37%(1)
NoE	1.69%(4)	2.34%(2)	0.00%(0)
Total	140	61	10
AGN	52.78±6.50%(66)	47.60±8.84%(29)	78.63±26.21%(9)

where D is the luminosity distance in Mpc and $S_{60\,\mu\rm m}$ and $S_{1.4\,\rm GHz}$ flux densities in units of Jy. The straight line is the best fitting line obtained by Yun et al. (2001) for an all-sky sample of infrared detected galaxies from *IRAS*,

$$\log L_{1.4\,\rm GHz} = (0.99 \pm 0.01) \log(L_{60\,\mu\rm m}/L_\odot) + (12.07 \pm 0.08)\,. \tag{12}$$

From these relations we find that the infrared–radio correlation for our subsamples follow the correlation for an all-sky sample of infrared detected galaxies from *IRAS* (Yun et al 2001) and Markarian galaxies studied by Luo & Wu (2005). The slight deviation in the PSC_SUB is not significant, being less than the scatter.

The parameter q is plotted for our subsample in Figure 11, following the formula (Condon et al. 1991):

$$q = \log\Big(\frac{2.58 S_{60\,\mu\rm m} + S_{100\,\mu\rm m}}{2.98\,\rm Jy}\Big) - \log\Big(\frac{S_{1.4\,\rm GHz}}{\rm Jy}\Big)\,. \tag{13}$$

In the figure the solid line at $q = 2.34$ marks the mean value obtained by Yun et al. (2001), and the top and bottom dotted lines are limits for three times FIR excess and radio excess from the mean, respectively. The radio excess objects are mainly Radio Loud (RL) AGNs (Roy & Norris 1997) that may have some complex mechanisms of energy generation (e.g. the jet emission).

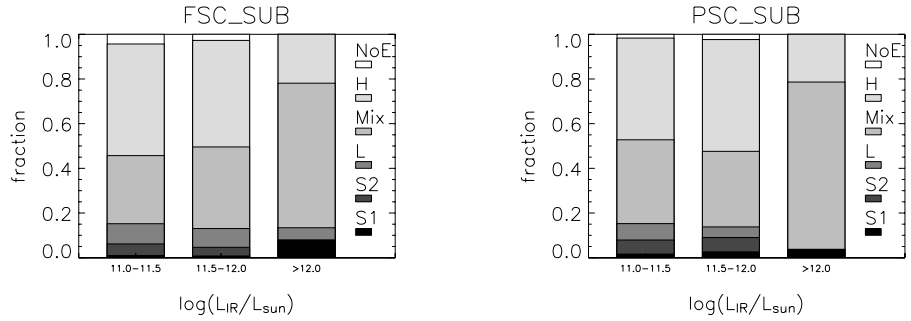

Fig. 9 Spectral type (Seyfert 1s (S1), Seyfert 2s (S2), LINERs (L), Mixed types (Mix), HIIs (H) and No apparent Emission lines (NoE), shaded from black to white) distribution in three infrared luminosity intervals. Left: the FSC subsample; Right: the PSC subsample.

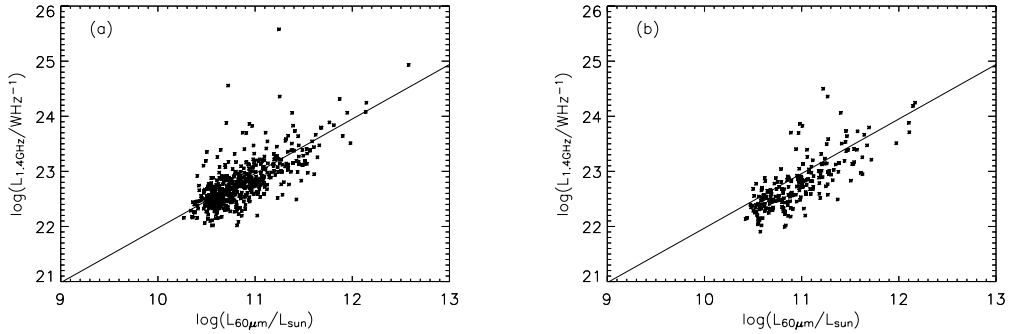

Fig. 10 Infrared (60 μm) – radio (1.4 GHz) correlation in the FSC subsample (a) and PSC subsample (b). The straight line is the best fit obtained by Yun et al. (2001) for an all-sky sample of infrared detected galaxies from *IRAS*.

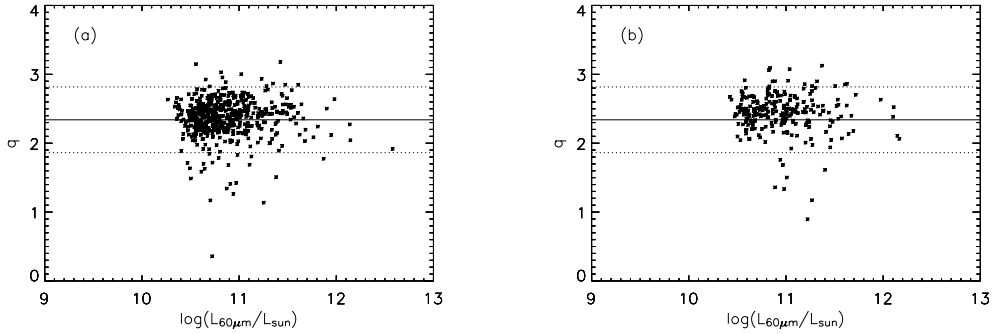

Fig. 11 Parameter q as function of the 60 μm luminosity for the FSC subsample (a), and the PSC subsample (b). The solid line is at $q = 2.34$ which is the mean value obtained by Yun et al. (2001). The top and bottom dotted lines are limits marking three times FIR excess and radio excess from the mean, respectively.

6 SUMMARY

In this paper we selected a sample of Luminous Infrared Galaxies based on cross-identification between the *IRAS* FSC and PSC data on one hand and the SDSS-DR2 on the other, and the results are given in a catalog. We use the "likelihood ratio" method to estimate the sample's reliability and for a high confidence subsample selection. Although the LR method has some problems and needs to be improved, it seems that it can be used as a stable and creditable sample selection method as judged by the analyses and comparisons in this work. From the statistical analyses (e.g., the redshift, $L_{\rm IR}$ and color distributions, the spectral types, and the radio to infrared correlations) we find that LIGs and ULIGs are quite different. We will perform further analyses in the future and attempt to reveal more about the LIGs, such as their morphologies and environments (Wang et al., in preparation), origins of the IR excess (Pasquali et al. 2005) and their star formation history. Some interesting subsamples like IR QSOs (Zheng et al. 2002; Hao et al. 2005) and RL AGNs (Best et al. 2005) will also be selected and analyzed for understanding connections between star formation and AGN activity. We will keep on seeking better statistical methods for huge astronomical data mining and analyses.

Acknowledgements We would like to thank Drs. J. Y. Wei, S. Mao, S. Komossa, J. Wang for advice and helpful discussions. We also thank the anonymous referee and the editor for very constructive comments and suggestions. This project is supported by the NSFC through Grants 10273012, 10333060 and 10473013, and NKBRSF G1999075404. 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/. 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, the Korean Scientist Group, 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, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington. This work also used the *IRAS* data from the PSC (IPAC 1986) and FSC (Moshir+ 1989), the FIRST data from the VLA-FIRST project and the emission line data from SDSS studies at MPA/JHU.

References

Abazajian K., Adelman-McCarthy J. K., Agüeros M. A. et al., 2004, AJ, 128, 502
Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5 (BPT)
Becker R. H., White R. L., Helfand D. J., 1995, ApJ, 450, 559
Best P. N., Kauffmann G., Heckman T. M., Ivezić Ž., 2005, MNRAS, 362, 9
Blanton M. R., Dalcanton J., Eisenstein D. et al., 2001, AJ, 121, 2358
Calzetti D., Armus L., Bohlin R. C. et al., 2000, ApJ, 533, 682
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Condon J. J., Anderson M. L., Helou G., 1991, ApJ, 376, 95
Condon J. J., 1992, ARA&A, 30, 575
Goto T., 2005a, MNRAS, 357, 937
Goto T., 2005b, MNRAS, 360, 322
Hao C. N., Xia X. Y., Mao S., Wu H., Deng Z. G., 2005, ApJ, 625, 78
Helou G., Soifer B. T., Rowan-Robinson M., 1985, ApJ, 298, L7
Helou G., Khan I., Malek L., Boehmer L., 1988, ApJS, 68, 151
Helou G., Bicay M. D., 1993, ApJ, 415, 93
Ivezić Ž., Menou K., Knapp G. R. et al., 2002, AJ, 124, 2364
Kauffmann G., Heckman T. M., White S. D. M. et al., 2003a, MNRAS, 341, 33
Kauffmann G., Heckman T. M., White S. D. M. et al., 2003b, MNRAS, 341, 54
Kauffmann G., Heckman T. M., Tremonti C. et al., 2003c, MNRAS, 346, 1055

Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, ApJ, 556, 121
Kim D. -C., Sanders D. B., Veilleux S., Mazzarella J. M., Soifer B. T., 1995, ApJS, 98, 129
Kim D. -C., Sanders D. B., 1998a, ApJS, 119, 41
Kim D. -C., Veilleux S., Sanders D. B., 1998b, ApJ, 508, 627
Kim D. -C., Veilleux S., Sanders D. B., 2002, ApJS, 143, 277
Lawrence A., Rowan-Robinson M., Leech K., Jones D. H., Wall J. V., 1989, MNRAS, 240, 329
Lonsdale C. J. et al., 1998, B. McLean, D. Golombek, J. Hayes, H. Payne, ed., IAU Colloq. 179, New Horizons from Multi-Wavelength Sky Surveys, Dordrecht: Kluwer, 450
Luo S.-G., Wu X.-B., 2005, ChJAA, 5, 448
Masci F. J., Condon J. J., Barlow T. A. et al., 2001, PASP, 113, 10
Moshir M., Kopan G., Conrow T. et al., 1990, IRAS Faint Source Catalogue, v. 2.0
Neugebauer G., Habing H. J., van Duinen R. et al., 1984, ApJ, 278, L1
Osterbrock D. E., 1989, Astrophysical of Gaseous Nebulae and Active Galactic Nuclei, Mill Valley CA: University Science Books
Osterbrock D. E., Pogge R. W., 1985, ApJ, 297, 166
Pasquali A., Kauffmann G., Heckman T. M., 2005, MNRAS, 361, 1121
Poggianti B. M., Wu H., 2000, ApJ, 529, 157
Roy A. L., Norris R. P., 1997, MNRAS, 289, 824
Rutledge R. E., Brunner R. J., Prince T. A., Lonsdale C., 2000, ApJS, 131, 335
Sanders D. B., Soifer B. T., Elias J. H. et al., 1988, ApJ, 325, 74
Sanders D. B., Mirabel I. F., 1996, ARA&A, 34, 749
Saunders W., Sutherland W. J., Maddox S. J. et al., 2000, MNRAS, 317, 55
Schmidt M., 1968, ApJ, 151, 393
Soifer B. T., Houck J. R., Neugebauer G., 1987a, ARA&A, 25, 187
Soifer B. T., Sanders D. B., Madore B. F. et al., 1987b, ApJ, 320, 238
Strateva I., Ivezić Ž., Knapp G. R. et al., 2001, AJ, 122, 1861
Sutherland W., Saunders W., 1992, MNRAS, 259, 413
Tremonti C. A., Heckman T. M., Kauffmann G. et al., 2004, ApJ, 613, 898
Veilleux S., Kim D. -C., Sanders D. B., Mazzarella J. M., Soifer B. T., 1995, ApJS, 98, 171
Veilleux S., Kim D. -C., Sanders D. B., 1999, ApJ, 522, 113
White R. L., Becker R. H., Helfand D. J., Gregg M. D., 1997, ApJ, 475, 479
Wu H., Zou Z. L., Xia X. Y., Deng Z. G., 1998a, A&AS, 127, 521
Wu H., Zou Z. L., Xia X. Y., Deng Z. G., 1998b, A&AS, 132, 181
Yang Y., Zabludoff A. I., Zaritsky D., Lauer T. R., Mihos J. C., 2004, ApJ, 607, 258
York D. G., Adelman J., Anderson J. E. Jr. et al., 2000, AJ, 120, 1579
Yun M. S., Reddy N. A., Condon J. J., 2001, ApJ, 554, 803
Zabludoff A. I., Zaritsky D., Lin Huan et al., 1996, ApJ, 466, 104
Zheng X. Z., Xia X. Y., Mao S., Wu H., Deng Z. G., 2002, AJ, 124, 18
Zheng X. Z., Hammer F., Flores H., Assémat, F., Pelat, D., 2004, A&A, 421, 847
Zou Z. L., Xia X. Y., Deng Z. G., Su H. J., 1991, MNRAS, 252, 593