SUBMITTED TO ApJ
Preprint typeset using LaTeX style emulateapj v. 10/10/03

THE CLUSTERING OF AGN IN THE SLOAN DIGITAL SKY SURVEY

DAVID A. WAKE[1], CHRISTOPHER J. MILLER[1], TIZIANA DI MATTEO[2], ROBERT C. NICHOL[1], ADRIAN POPE[3], ALEXANDER S. SZALAY[3], ALEXANDER GRAY[4], DONALD P. SCHNEIDER[5], DONALD G. YORK[6]

Submitted to ApJ

ABSTRACT

We present the two–point correlation function (2PCF) for narrow-line active galactic nuclei (AGN) selected within the First Data Release of the Sloan Digital Sky Survey. Using a sample of 13605 AGN in the redshift range $0.055 < z < 0.2$, we find that the AGN auto–correlation function is consistent with the observed galaxy auto–correlation function on scales $0.2h^{-1}$Mpc to $> 100h^{-1}$Mpc. The AGN hosts trace an intermediate population of galaxies and are not detected in the bluest (youngest) disk–dominated galaxies or many of the reddest (oldest) galaxies. We show that the AGN 2PCF is dependent on the luminosity of the narrow [OIII] emission line ($L_{[OIII]}$), with low $L_{[OIII]}$ AGN having a higher clustering amplitude than high $L_{[OIII]}$ AGN. This is consistent with lower activity AGN residing in more massive galaxies than higher activity AGN, and $L_{[OIII]}$ providing a good indicator of the fueling rate. Using an AGN model embedded in cosmological simulations, we show that AGN hosted by $\sim 10^{12}$ M$_\odot$ dark matter halos have a 2PCF that matches that of the observed sample. This mass scale implies a mean black hole mass for the sample of $M_{BH} \sim 10^8$ M$_\odot$.

Subject headings: galaxies: active — galaxies: active — galaxies: statistics

1. INTRODUCTION

The clustering of galaxies as a function of their properties provides important constraints on models of galaxy formation and evolution. Such clustering is often measured using the two–point auto–correlation function (2PCF; see Peebles 1980). In hierarchical models of structure formation, the amplitude of the 2PCF depends upon the mass of the dark matter halo (*i.e.,* more massive halos are clustered more strongly; Kaiser 1986), while the shape of the 2PCF can depend upon the details of how galaxies reside in those dark matter halos (Zehavi et al. 2004). For example, the amplitude and slope of the 2PCF is lower for blue galaxies than for galaxies with the reddest rest-frame colors (e.g. Zehavi et al. 2002).

In this letter, we continue our study of the relation between the environment of galaxies in the Sloan Digital Sky Survey (SDSS; York et al. 2000) and their observed physical properties (see Gómez et al. 2003; Miller et al. 2003; Balogh et al. 2004). In particular, we present the redshift–space 2PCF for a subset of SDSS galaxies spectroscopically classified as narrow-line active galactic nuclei (AGN; Miller et al. 2003). Our analysis has two advantages over previous measurements of the AGN 2PCF: larger sample size (in number and area), and a homogeneous selection criteria (compare to Table 1 of Brown et al. (2001) for previous AGN 2PCF measurements). In addition, the data are now large enough to study volume-limited subsamples as well as to study how AGN or AGN host galaxy properties affect the 2PCF.

[1] Dept. of Physics, Carnegie Mellon Univ., Pittsburgh, PA 15213
[2] Max Planck Institute for Astrophysics Garching, D-85741 Garching, Germany
[3] Dept. of Physics and Astronomy, Johns Hopkins Univ., Baltimore, MD
[4] Dept. of Computer Science, Carnegie Mellon Univ., Pittsburgh, PA 15217
[5] Dept. of Astronomy and Astrophysics, Pennsylvania State Univ., University Park, PA 16802
[6] Dept. of Astronomy and Astrophysics and Enrico Fermi Institute, The University of Chicago, Chicago, IL 60637

2. DATA

We use the main galaxy sample data (Strauss et al. 2002) from the First Data Release (DR1) of the SDSS (see Abazajian et al. 2003). To select AGN, we have used the methodology presented in Section 2.1 of Miller et al. (2003), where the AGN are classified using emission-line flux ratios (log([OIII]/Hβ) versus log([NII]/Hα) (see Kewley et al. 2001)) or simply log([NII]/Hα)> -0.2, if [OIII] or Hβ are not measured (see also Carter et al. 2001; Brinchmann et al. 2004). We remove all galaxies from areas with high seeing values ($> 2''$) and r band Galactic extinction > 0.4 magnitudes. These restrictions produce a sample of 72455 SDSS DR1 galaxies within $0.055 \leq z \leq 0.2$, from which we classify 13605 galaxies as AGN. This percentage of AGN (18%) is consistent with the findings of Miller et al. (2003) and Brinchmann et al. (2004). We discuss implications of our classifications in Section 4.1

3. AGN AND GALAXY CORRELATION FUNCTIONS

We account for the survey geometry (or mask) by constructing random catalogs that match both the survey angular and radial selection functions. We first construct a random catalog that has the same angular mask as the real data. We then construct the radial selection function by heavily smoothing the observed redshift distributions with a Gaussian of width $z = 0.025$. These smoothed redshift distributions are used to randomly assign redshifts to the data points in our random catalogs, which are ten times larger than the real datasets. We calculate the 2PCF using the Landy & Szalay (1993) estimator and estimate the covariance using the jack-knife resampling technique (e.g. Lupton 1993; Zehavi et al. 2002). We use $H_0 = 100$km s^{-1} Mpc^{-1}, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$.

In Figure 1, we show the redshift–space 2PCF for both the AGN (stars) and all galaxies (filled dots) samples discussed in Section 2. We also show the SDSS redshift-space 2PCF of Zehavi et al. (2002). As expected, our galaxy 2PCF agrees with Zehavi et al. (2002), except on

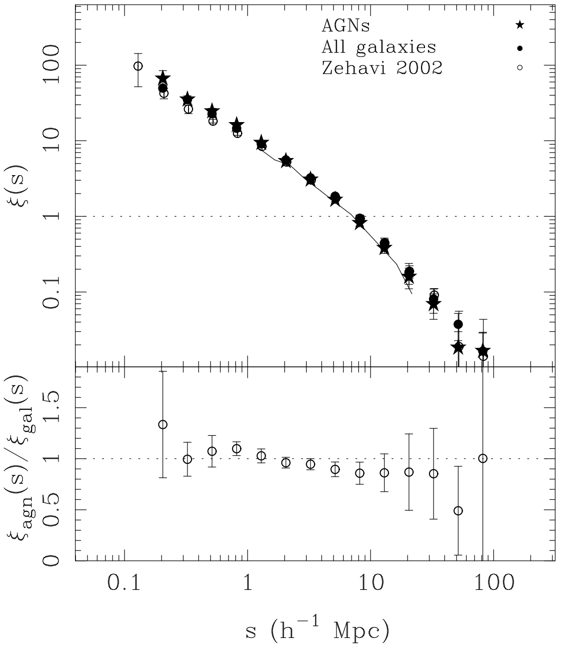

FIG. 1.— (Top) The 2PCF for AGN and all SDSS galaxies. We also show the 2PCF of Zehavi et al. (2002). The solid line is the model from Section 4.2. (Bottom) The ratio of the AGN and galaxy 2PCFs.

small scales due to incompleteness from fiber collisions (Blanton et al. 2003a), since Zehavi et al. attempt to correct for these collisions. We note that since the fiber collisions are uniformly distributed over the survey (see Blanton et al. 2003a), they affect both the galaxy and AGN 2PCF in the same way. Thus we study only relative differences between 2PCFs in this work. A statistical analysis using the sum-squared covariances between the 2PCFs for all galaxies and AGN gives $\chi^2 = 13$ with 9 degrees-of-freedom for pair separations greater than $\sim 1h^{-1}$Mpc, *i.e.*, there is no significant difference between the AGN and galaxy 2PCF. We show the of the ratio of the two 2PCFs in the bottom of Figure 1. The weighted mean ratio is $\xi_{agn}/\xi_{gal} = 0.974 \pm 0.026$.

We show in Figure 2 the ratio of the AGN–galaxy cross–correlation function (Croft et al. 1999; Croom et al. 2003) to the normal galaxy 2PCF. The weighted mean ratio between these two functions is $\xi_{agn-gal}/\xi_{gal-gal} = 0.922 \pm 0.028$. Within the one sigma uncertainties, this is consistent with Croom et al. (2003) who demonstrate that, for $z < 0.3$ quasars, the ratio of the quasar–galaxy cross correlation function is $\xi_{QSO-gal}/\xi_{gal-gal} = 0.97 \pm 0.05$.

Finally, we measure the 2PCF as a function of AGN activity as measured by the luminosity of the forbidden [OIII] narrow emission line ($L_{[OIII]}$). This line is proposed to be only weakly affected by any residual star–formation (see Kauffmann et al. 2003) and should provide a measure of the AGN activity. However, the exact connection between the [OIII] strength and AGN activity is not well established in narrow-line systems (see Nelson 2000; Boroson 2002; Mathur 2000; Mathur et al. 2001; Grupe & Mathur 2004). We created two sub-samples of AGN, one containing the highest third of $L_{[OIII]}$ ($> 4.84 \times 10^{44}$ ergs^{-1}) of the distribution and one with the lowest

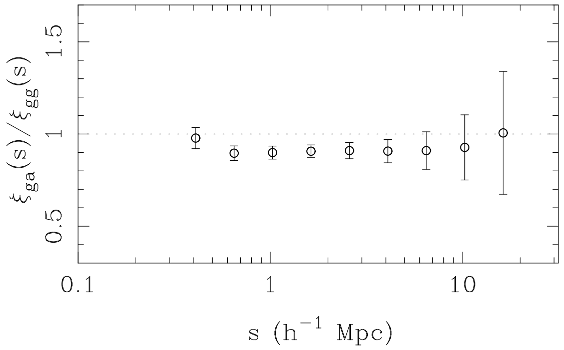

FIG. 2.— The ratio of the AGN–galaxy cross–correlation function to the galaxy–galaxy auto correlation function.

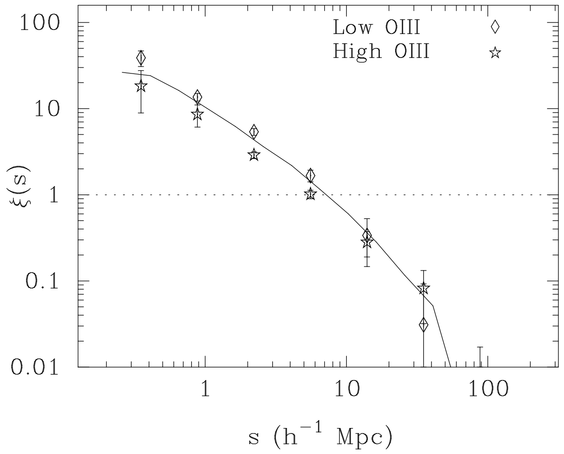

FIG. 3.— The AGN 2PCF as a function of the luminosity of the [OIII] narrow emission line. We show the 2PCF for AGN in the top and bottom third of the [OIII] luminosity distribution. The solid line is for all SDSS galaxies taken from Figure 1.

third of $L_{[OIII]}$ ($< 1.29 \times 10^{44}$ ergs^{-1}). In order to minimize any selection bias (*e.g.* aperture effects), we construct a pseudo-volume limited sample by restricting the AGN sample to a redshift range of $0.06 < z < 0.085$ and to a k-corrected (Blanton et al. 2003b) absolute magnitude limit of $M_r < -19.8$ (see Gómez et al. 2003; Balogh et al. 2004). This provides a sample of 2457 AGN. We find that the distributions of host absolute magnitudes and redshifts are identical for the low and high $L_{[OIII]}$ samples and to the entire galaxy sample.

In Figure 3, we present the AGN 2PCF as a function of $L_{[OIII]}$. We see a noticeable difference in the amplitude of clustering, with the lower luminosity AGN having a stronger clustering amplitude. As in Section 3, we calculate the χ^2 difference between the high and low $L_{[OIII]}$ sub-samples and find $\chi^2 = 40$ with 5 degrees-of-freedom. Therefore, the 2PCFs for the high and low $L_{[OIII]}$ sub-samples are different at the $> 5\sigma$ level. We also see a similar difference in the clustering amplitudes if we split the AGN sample as a function of the width of the [OIII] emission–line.

4. DISCUSSION

4.1. *The Properties of AGN Host Galaxies*

Using our magnitude and volume-limited sample, we find that the distributions of the AGN host galaxy prop-

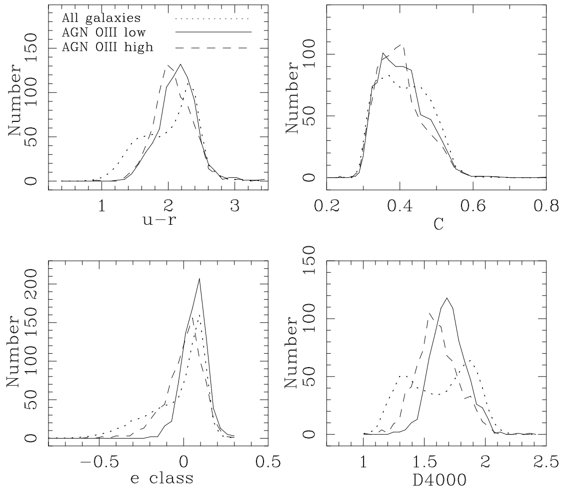

FIG. 4.— The distributions of galaxy properties (described in the text) for our volume–limited sample. All distributions have been re–normalized to have the same total number.

erties are different from those of all galaxies (Figure 4). For example, the concentration index (C) (Shimasaku et al. 2001) and e-class (Stoughton et al. 2002; Connolly & Szalay 1998) for our AGN sample appear to be preferentially missing the bluest (in e-class), disk–dominated galaxies (see also Figure 12 in Miller et al. 2003). The most striking difference is for the D4000, a measure of the 4000Åbreak, and the $u-r$ color, where the AGN distributions do not exhibit the bi-modal shape seen for all galaxies.

We are assuming that these properties reflect that of the host galaxies and are not being affected by the AGN light. Both Schmitt, Storchi-Bergmann, & Fernandes (1999) and Kauffmann et al. (2003) find that the AGN contribution to the total luminosity in narrow-line AGN like those studied here rarely exceeds 5%. We estimated the AGN contribution using the ratio of the total flux over the flux within the 3″fiber. We compared these ratios for hosts without AGN to our AGN hosts as a function of $L_{[OIII]}$. In our highest $L_{[OIII]}$, sample, the AGN contribute on average $< 6\%$ to the total light, in agreement with Schmitt, Storchi-Bergmann, & Fernandes (1999) and Kauffmann et al. (2003).

One possible explanation for the difference between the AGN and all galaxy distributions is our exclusion of the broad-line QSO. To investigate this, we have identified 94 broad-line QSOs within $0.06 \leq z \leq 0.085$ and brighter than M_r = -19.8. These QSO host galaxies have a wide variety of morphologies, however, 20% lack an obvious host galaxy (i.e., they appear point-like in the SDSS), and so any k-corrections on this population could be inaccurate. However, the observed-frame colors of all of the QSOs are on average bluer than our AGN sample. However, this is expected if the QSO component is dominating over the light from the host galaxy. Even so, the total number of these broad-line QSOs is simply too small to accommodate the missing blue galaxy populations discussed above.

Another possibility comes from our AGN selection criteria. In particular, our signal–to–noise limit on the SDSS spectra could preferentially exclude the lowest luminosity AGN, as they could be buried in strongly star forming bulges or accreting at a very low rate in the reddest, oldest galaxies. Dust obscuration could also significantly affect our detection of AGN, especially for the strongly star-forming (bluest) galaxies (Hopkins et al. 2003). For instance, roughly $\sim 30\%$ of our galaxies are emission-line galaxies that could not be classified as either star-forming or AGN, and could be obscured AGN. Miller et al. (2003) attempted to statistically model this population using colors and noted that there was a significant red population, most likely AGN. Using this model, we have classified these unidentified emission-line galaxies as AGN and recalculated the 2PCF. We find no statistical difference after the inclusion of these model-dependent AGN classifications.

In summary, the exclusion of the QSOs does not explain why our AGN sample is lacking the bi-modal color distribution of the whole galaxy sample. Likewise, while we are certainly missing some AGN in the unidentified emission-line galaxy population, our measured 2PCF is not altered after we attempt to include them. These issues could be addressed in a more detailed way through a multi–wavelength study of these unidentified emission-line objects. Therefore, as suggested in Miller et al. (2003), our AGN sample appears to be an unbiased tracer, with respect to mass of the whole galaxy population, for the large-scale structure in the local Universe.

Given that the typical AGN host properties are not a random sub-sample of all galaxies, it is somewhat surprising that they should cluster the same way as all galaxies. For example, it has been shown that the SDSS 2PCF is a strong function of both the color and luminosity of galaxies (Zehavi et al. 2002), indicating that the bluest, youngest galaxies preferentially populate the lowest density regions in the Universe (Gómez et al. 2003; Balogh et al. 2004), while the reddest, oldest galaxies preferentially live in the densest regions (*e.g.*, cores of rich clusters). Removing these two tails of the distribution could result in an intermediate population, selected here as our AGN hosts, that cluster the same way as the whole sample. We tested this scenario using cosmological simulations (see next Section) and found that this would be plausible if the mean mass scale of the two samples is the same.

4.2. *The Mass Scale Selected by the AGN*

We have used the analytical model of Di Matteo et al. (2003b) to make a prediction for the AGN 2PCF. In this model, cosmological hydrodynamical simulations (Springel & Hernquist 2003a,b) were used to link the growth and activity of central black holes in galaxies to the formation of spheroids in galaxy halos In the prescription used for black hole growth in the simulations, it is assumed that the black hole fueling rate is regulated by star formation in the gas. This simple assumption was shown to explain the observed $M-\sigma$ relation (Ferrarese & Merritt 2000; Gebhardt et al. 2000) and the broad properties of the AGN luminosity function (for an assumed quasar lifetime). We use this model to construct a mock AGN catalog and deduce that the minimum dark matter halo mass ($M_{\rm min}$) in the simulations that best matches the observed space density of SDSS AGN ($n \sim 1.5 \times 10^{-3}$ Mpc^{-3}) is $M_{\rm min} \geq 2 \times 10^{12}$ M$_{\odot}$. This is representative of low redshift L^{*} galaxies, which is the bulk of our AGN sample. Not surprisingly, the 2PCF for simulated dark matter halos with masses greater than

$M_{\rm min}$ (shown in Figure 1) agrees well with the observed 2PCF (also based on $\sim L^*$ and brighter galaxies). As mentioned in the last section, we expect the AGN clustering amplitude to match that of the entire galaxy population, when the mean mass for the two samples were similar, which is what we infer from the simulations.

By relating the dark matter halo to the black hole mass (according to Eqn. 8 of Di Matteo et al. 2003b; see also the observed correlation by Ferrarese 2001 and Baes et al. 2003), we deduce a mean black hole mass of our sample of SDSS AGN, $M_{BH} \sim 10^8$ $M_\odot$.

4.3. *Clustering and AGN Activity*

In Figure 3, we found that the 2PCF for the lowest $L_{[OIII]}$ AGN in our sample appear to have a higher clustering amplitude than the highest $L_{[OIII]}$ AGN. We tested whether this amplitude difference is a result of the differing AGN host galaxy distributions. We randomly constructed two AGN sub-samples that possess the same D4000 distributions as shown in Figure 4 for the low and high $L_{[OIII]}$ samples, with no regard to the $L_{[OIII]}$. We found no difference in their respective 2PCFs. We repeated the test for $u-r$ color, e-class and concentration index and again find no difference. These tests demonstrate that the difference seen in the clustering strengths between the low and high $L_{[OIII]}$ samples is driven by the [OIII] emission line and not by the underlying galaxy properties in Figure 4. As an additional test, we split our AGN sample into highest and lowest thirds of their D4000s, $u-r$ color, e-class, and concentration indices, regardless of their $L_{[OIII]}$. In most cases, we do in fact see clustering differences in these subsamples. For instance, the high D4000 sample is more strongly clustered than the low D4000 AGN sample. Likewise, the redder AGN are more clustered than the bluer AGN. However, the difference in clustering amplitude is strongest when the AGN are split by $L_{[OIII]}$.

In hierarchical models of structure formation, more massive dark matter halos are more strongly clustered. Therefore, the fact that the lower $L_{[OIII]}$ AGN subsample has a higher clustering amplitude indicates that the host dark matter haloes of these AGN must be preferentially more massive then the weaker $L_{[OIII]}$ AGN. Futhermore, $L_{[OIII]}$ delineates the high and low mass halos better than other host galaxy properties (like D4000 or color). If as expected the mass of the black hole correlates with the halo mass, the weaker $L_{[OIII]}$ AGN must have larger black holes, and therefore a low $L_{[OIII]}$ can only be caused by a low fueling rate. These observations are in accordance with studies of nearby massive ellipticals, which are known to host the largest black hole masses (consistent with the $M-\sigma$ relation Gebhardt et al. 2000; Ferrarese & Merritt 2000), but which typically display the weakest AGN (Ho, Filippenko, & Sargent 1997; Di Matteo et al. 1999, 2003a, LINERs). Conversely, the high $L_{[OIII]}$ AGN have a lower clustering amplitude consistent with them being in lower mass dark matter halos (hence having lower central black holes) but accreting at a high rate.

We thank Volker Springel and Lars Hernquist for their Hydro-simulations which were performed at the Center for Parallel Astrophysical Computing at the Harvard-Smithsonian Center for Astrophysics and Jon Loveday, David Weinberg and Michael Vogeley for useful discussions. 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/.

REFERENCES

Abazajian, K. et al. 2003, AJ, 126, 2081
Baes, M., Buyle, P., Hau, G. K. T., & Dejonghe, H. 2003, MNRAS, 341, L44
Balogh, M. et al. 2004, submitted MNRAS, astro-ph/0311379
Blanton, M. R., Lin, H., Lupton, R. H., Maley, F. M., Young, N., Zehavi, I., & Loveday, J. 2003a, AJ, 125, 2276
Blanton, M. R. et al. 2003b, AJ, 125, 2348
Boroson, T. A. 2002, ApJ, 565, 78
Brinchmann, M. et al. 2004, submitted MNRAS, astro-ph/0311060
Brown, M. J. I., Boyle, B. J., & Webster, R. L. 2001, AJ, 122, 26
Carter, B. J., Fabricant, D. G., Geller, M. J., Kurtz, M. J., & McLean, B. 2001, ApJ, 559, 606
Connolly, A. J. & Szalay, A. S. 1998, IAU Symp. 179: New Horizons from Multi-Wavelength Sky Surveys, 179, 376
Croft, R. A. C., Dalton, G. B., & Efstathiou, G. 1999, MNRAS, 305, 547
Croom, S. et al. 2003, astro-ph/0310533
Di Matteo, T., Fabian, A. C., Rees, M. J., Carilli, C. L., & Ivison, R. J. 1999, MNRAS, 305, 492
Di Matteo, T., Allen, S. W., Fabian, A. C., Wilson, A. S., & Young, A. J. 2003a, ApJ, 582, 133
Di Matteo, T., Croft, R. A. C., Springel, V., & Hernquist, L. 2003b, ApJ, 593, 56
Ferrarese, L. & Merritt, D. 2000, ApJ, 539, L9
Ferrarese, L., Pogge, R. W., Peterson, B. M., Merritt, D., Wandel, A., & Joseph, C. L. 2001, ApJ, 555, L79
Gebhardt, K. et al. 2000, ApJ, 539, L13
Gómez, P. L. et al. 2003, ApJ, 584, 210
Grupe, D. & Mathur, S. 2004, submitted ApJ, astro-ph/0312390
Ho, L. C., Filippenko, A. V., & Sargent, W. L. W. 1997, ApJ, 487, 568
Hopkins, A. M. et al. 2003, ApJ, 599, 971
Kaiser, N. 1986, MNRAS, 222, 323
Kauffmann, G. et al. 2003, MNRAS, 346, 1055
Kauffmann, G. et al. 2004, MNRAS, submitted (astro-ph/0402030)
Kewley, L. J., Heisler, C. A., Dopita, M. A., & Lumsden, S. 2001, ApJS, 132, 37
Landy, S. D. & Szalay, A. S. 1993, ApJ, 412, 64
Lupton, R. 1993, Princeton, N.J.: Princeton University Press, —c1993,
Mathur, S. 2000, MNRAS, 314, L17
Mathur, S., Kuraszkiewicz, J., & Czerny, B. 2001, New Astronomy, 6, 321
Miller, C. J., Nichol, R. C., Gómez, P. L., Hopkins, A. M., & Bernardi, M. 2003, ApJ, 597, 142
Nelson, C. H. 2000, ApJ, 544, L91
Peebles, P. J. E. 1980, Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p.,
Schmitt, H. R., Storchi-Bergmann, T., & Fernandes, R. C. 1999, MNRAS, 303, 173
Shimasaku, K. et al. 2001, AJ, 122, 1238
Springel, V. & Hernquist, L. 2003a, MNRAS, 339, 289
Springel, V. & Hernquist, L. 2003b, MNRAS, 339, 312
Stoughton, C. et al. 2002, AJ, 123, 485
Strauss, M. A. et al. 2002, AJ, 124, 1810
York, D. G. et al. 2000, AJ, 120, 1579
Zehavi, I. et al. 2002, ApJ, 571, 172
Zehavi, I. et al. 2004, submitted ApJ, astro-ph/0301280