The Astrophysical Journal, 610:L85–L88, 2004 August 1

THE CLUSTERING OF ACTIVE GALACTIC NUCLEI 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] and Donald G. York[6]

Received 2004 March 31; accepted 2004 June 15; published 2004 June 29

ABSTRACT

We present the two-point correlation function (2PCF) of narrow-line active galactic nuclei (AGNs) selected within the First Data Release of the Sloan Digital Sky Survey. Using a sample of 13,605 AGNs in the redshift range $0.055 < z < 0.2$, we find that the AGN autocorrelation function is consistent with the observed galaxy autocorrelation function on scales from 0.2 to greater than 100 h^{-1} Mpc. The AGN hosts trace an intermediate population of galaxies and are not detected in either 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 [O III] emission line ($L_{[\mathrm{O\,III}]}$), with low $L_{[\mathrm{O\,III}]}$ AGNs having a higher clustering amplitude than high $L_{[\mathrm{O\,III}]}$ AGNs. This is consistent with lower activity AGNs residing in more massive galaxies than higher activity AGNs, and $L_{[\mathrm{O\,III}]}$ providing a good indicator of the fueling rate. Using a model relating halo mass to black hole mass in cosmological simulations, we show that AGNs 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 $M_{\mathrm{BH}} \sim 10^{8}\ M_{\odot}$.

Subject headings: 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 correlation function (2PCF; see Peebles 1980). In hierarchical models of structure formation, the amplitude of the 2PCF depends on the mass of the dark matter halos (i.e., more massive halos are clustered more strongly; Kaiser 1986), while the shape of the 2PCF can depend on 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 colors (e.g., Davis & Geller 1976; 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, and 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 (AGNs; Miller et al. 2003). Our analysis has two advantages over previous measurements of the AGN 2PCF: a 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 both volume-limited subsamples as well as how AGNs or AGN host galaxy properties affect the 2PCF.

[1] Department of Physics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213.
[2] Max Planck Institute for Astrophysics, Karl-Schwarzschild-Strasse 1, Postfach 1317, D-85741 Garching, Germany.
[3] Department of Physics and Astronomy, Johns Hopkins University, Charles and 34th Street, Bloomberg Center, Baltimore, MD 21218-2686.
[4] Department of Computer Science, Carnegie Mellon University, 4212 Wean Hall, Pittsburgh, PA 15213.
[5] Department of Astronomy and Astrophysics, Pennsylvania State University, 525 Davey Laboratory, University Park, PA 16802.
[6] Department of Astronomy and Astrophysics and Enrico Fermi Institute, University of Chicago, 5640 South Ellis Avenue, 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 AGNs, we have used the methodology presented in § 2.1 of Miller et al. (2003), where the AGNs are classified using the emission-line flux ratios log ([O III]/Hβ) versus log ([N II]/Hα) (see Kewley et al. 2001) or simply log ([N II]/Hα) > -0.2, if [O III] or Hβ are not measured (see also Carter et al. 2001 and Brinchmann et al. 2004). We remove all galaxies from areas with high seeing values (>2″) and *r*-band Galactic extinction greater than 0.4 mag. These restrictions produce a sample of 72,455 SDSS DR1 galaxies within $0.055 \le z \le 0.2$, from which we classify 13,605 galaxies as AGNs. This fraction of AGNs (18%) is consistent with the findings of Miller et al. (2003) and Brinchmann et al. (2004). We discuss the implications of our classifications in § 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 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 10 times larger than the real data sets. We note that our conclusions are robust to the choice of width for the smoothing kernel.

We calculate the 2PCF using the Landy & Szalay (1993) estimator and estimate the covariance using the jackknife resampling technique (e.g., Lupton 1993; Zehavi et al. 2002), splitting the angular mask of our data into 32 subsections of $\simeq 10^{\circ} \times 5^{\circ}$ (or $\simeq 60 \times 30\ h^{-1}$ Mpc at $z = 0.1$). 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 AGNs (*filled stars*) and all of the galaxies (*filled circles*) discussed in § 2. We also show the SDSS redshift-space 2PCF of Zehavi et al. (2002). As expected, our galaxy 2PCF agrees with that of Zehavi et al. (2002), except on small scales because of incom-

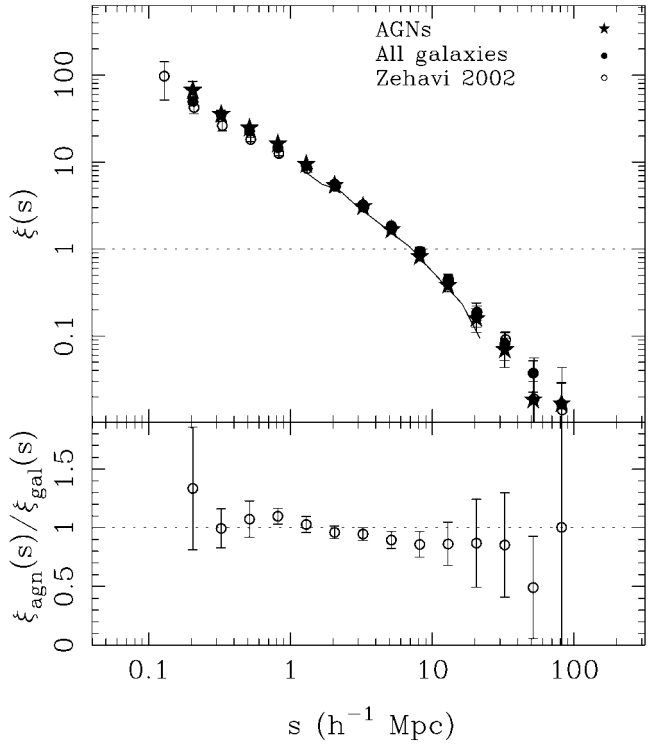

FIG. 1.—*Top*: 2PCF for AGNs and all SDSS galaxies. We also show the 2PCF of Zehavi et al. (2002). The solid line is the model from § 4.2. *Bottom*: Ratio of AGN to galaxy 2PCFs.

pleteness 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 2PCF and the AGN 2PCF in the same way. Thus, here we study only the relative differences between the 2PCFs. We measure a simple χ^2 statistic between the two samples and account for errors on both 2PCFs by combining their individual covariances. Specifically, we take the square root of the sum of the squared covariances, which accounts for correlations between data points of different separations in both data sets. We find $\chi^2 = 13$ with 9 degrees of freedom (dof) for pair separations greater than $\sim 1\ h^{-1}$ Mpc; i.e., there is no significant difference between the AGN 2PCF and the galaxy 2PCF. We show the ratio of the two 2PCFs in the bottom of Figure 1. The weighted mean ratio is $\xi_{\rm agn}/\xi_{\rm gal} = 0.974 \pm 0.026$.

We show in Figure 2 the ratio of the AGN-galaxy cross-correlation function (see also Croft et al. 1999 and Croom et al. 2003) to the normal galaxy 2PCF, again for the samples defined in § 2. The weighted mean ratio between these two functions is $\xi_{\rm agn\text{-}gal}/\xi_{\rm gal\text{-}gal} = 0.922 \pm 0.028$. Within the 1 σ 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 $\xi_{\rm QSO\text{-}gal}/\xi_{\rm gal\text{-}gal} = 0.97 \pm 0.05$.

Finally, we measure the 2PCF as a function of AGN activity as measured by the luminosity of the forbidden [O III] narrow emission line ($L_{\rm [O\ III]}$). This line is proposed to be only weakly affected by any residual star formation (see Kauffmann et al. 2003). However, the exact connection between the [O III] strength and AGN activity is not well established in narrow-line systems (see Nelson 2000, Boroson 2002, Mathur 2000, Mathur et al. 2001, and Grupe & Mathur 2004). We created two subsamples of AGNs, one containing the highest third of the $L_{\rm [O\ III]}$ distribution ($>4.84 \times 10^{44}$ ergs s^{-1}) and one with the lowest third ($<1.29 \times 10^{44}$ ergs s^{-1}). In order to minimize any selection bias (e.g., Malmquist bias), 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 and Balogh et al. 2004). This provides a sample of 2457 AGNs. We find that the distributions of host galaxy absolute magnitudes and redshifts are identical for the low and high $L_{\rm [O\ III]}$ samples and for the entire galaxy sample.

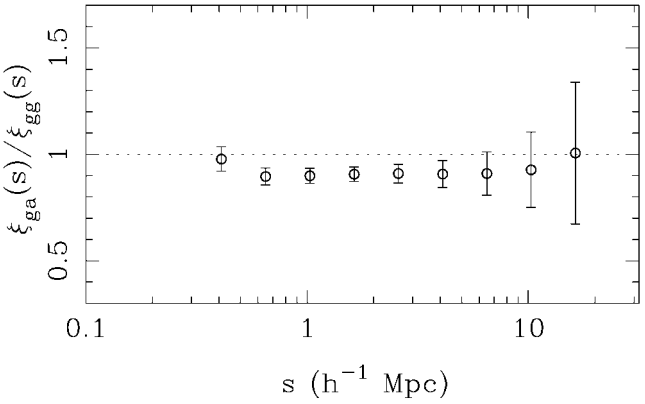

FIG. 2.—Ratio of the AGN-galaxy cross-correlation function to the galaxy-galaxy autocorrelation function.

In Figure 3, we present the AGN 2PCF as a function of $L_{\rm [O\ III]}$. We see a noticeable difference in the amplitude of clustering, with the lower luminosity AGNs having a stronger clustering amplitude. We calculate the χ^2 difference between the high and low $L_{\rm [O\ III]}$ subsamples and find $\chi^2 = 40$ with 5 dof. Therefore, the 2PCFs for the high and low $L_{\rm [O\ III]}$ subsamples are different at the greater than 5 σ 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 [O III] emission line.

4. DISCUSSION

4.1. *The Properties of AGN Host Galaxies*

Throughout this Letter we assume that the measured galaxy properties reflect those of the host galaxies and are not sig-

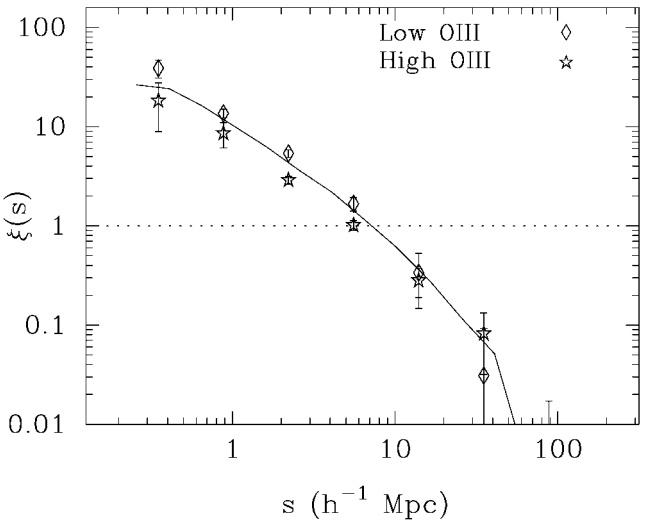

FIG. 3.—AGN 2PCF as a function of $L_{\rm [O\ III]}$. We show the 2PCF for AGNs in the top and bottom third of the $L_{\rm [O\ III]}$ distribution. The solid line is for all SDSS galaxies taken from Fig. 1.

nificantly affected by the AGN light. Both Schmitt et al. (1999) and Kauffmann et al. (2003) find that the AGN contribution to the total luminosity in narrow-line AGNs like those studied here rarely exceeds 5%. We measure the AGN contribution using the ratio of the total flux over the flux within the 3″ fiber and find that the AGNs contribute on average less than 6% of the total light.

Using our magnitude- and volume-limited sample, we find that the distributions of the AGN host galaxy properties are different from those of all galaxies (Fig. 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 Fig. 12 in Miller et al. 2003). The most striking difference is for the D4000 index, a measure of the 4000 Å break, and the $u - r$ color, where the AGN distributions do not exhibit the bimodal shape seen for all galaxies.

One possible explanation for the difference between the AGN distributions and the galaxy distributions is our exclusion of the broad-line QSOs. To investigate this, we have used the 94 broad-line SDSS 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, although 20% lack an obvious host galaxy (i.e., they appear pointlike 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, as expected if the QSO component dominates 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 host galaxy populations discussed above.

Another possibility comes from our AGN selection criteria. In particular, our signal-to-noise ratio limit on the SDSS spectra could preferentially exclude the lowest luminosity AGNs, as they could be either 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 AGNs, especially for the strongly star-forming (bluest) galaxies (Hopkins et al. 2003). For instance, ~30% of our galaxies have emission lines but could not be classified as either star-forming or AGNs, and could be obscured AGNs. Miller et al. (2003) attempted to statistically model this population using colors, and they noted that there was a significant red population, which was most likely AGNs. Using this model, we have classified these unidentified emission-line galaxies (ELUs) as AGNs. This has the effect of adding mainly red (but some blue) galaxies to the histograms in Figure 4, although they are still dominated by galaxies intermediate between blue and red (low and high D4000). We recalculate the 2PCF, including these model-dependent AGN classifications, and find no statistical difference from the AGN 2PCF that ignores the ELUs. We have not attempted to subtract off the stellar components of the SDSS spectra (see, e.g., Hao & Strauss 2004), and so we cannot rigorously address those galaxies that have both star-forming and AGN components. However, as noted in Miller et al. (2003), the fraction of late-type (e.g., spiral) morphologically classified galaxies harboring an AGN is ~20%, which is similar to that found by Ho (2004) who does subtract off stellar templates. The consistency between Ho (2004) and Miller et al. (2003) suggests that there are no large numbers of AGNs in strongly star-forming galaxies that we fail to detect.

In summary, the exclusion of the QSOs does not explain why our AGN sample is lacking the bimodal color distribution of the whole galaxy sample. Likewise, while we are certainly missing some AGNs 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 multiwavelength 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 the mass of the whole galaxy population, for the large-scale structure in the local universe.

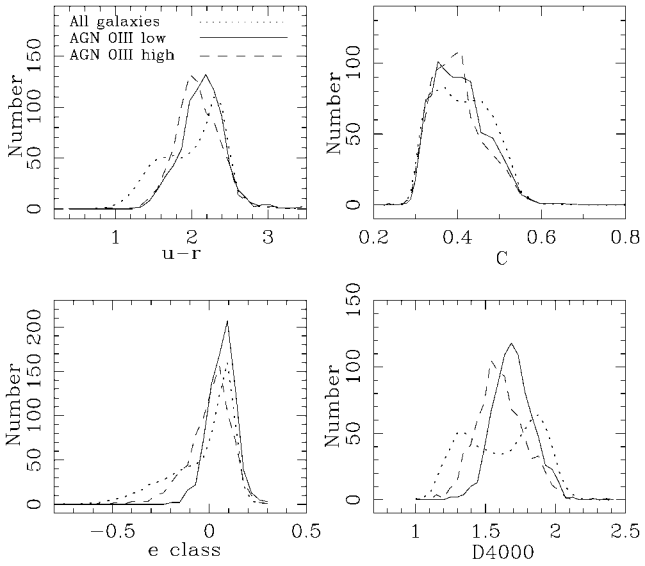

FIG. 4.—Distributions of galaxy properties (described in the text) for our volume-limited sample. All distributions have been renormalized to have the same total number.

Given that the typical AGN host properties are not a random subsample 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 2PCF is a strong function of both the color and the luminosity of galaxies (Davis & Geller 1976; Hamilton 1988; Zehavi et al. 2002), indicating that the bluest, youngest galaxies preferentially populate the lowest density regions in the universe, while the reddest, oldest galaxies preferentially live in the densest regions (e.g., Oemler 1974; Dressler 1980; Gómez et al. 2003; Balogh et al. 2004). Removing these two tails of the distribution could result in an intermediate population that clusters the same way as the whole sample. Assuming that the existence of an AGN is independent of environment (see Miller et al. 2003), one can conclude that the mean mass of the AGN dark matter halos must be the same as the mean for all galaxies (see § 4.2).

4.2. *The Mass Scale Selected by the AGNs*

We have used the model of Di Matteo et al. (2003b) to make a prediction for the AGN 2PCF. In this model, cosmological hydrodynamical simulations (Springel & Hernquist 2003a, 2003b) 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 both the observed M-σ 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 AGNs ($n \sim 1.5 \times 10^{-3}$ Mpc^{-3}) is $M_{\rm min} \geq 2 \times 10^{12}\ M_{\odot}$. This is representative of low-redshift L^* galaxies, which provide the bulk

of our AGN sample. Not surprisingly, the 2PCF for simulated dark matter halos with masses greater than $M_{\rm min}$ (shown in Fig. 1) agrees well with the observed 2PCF (also based on $\sim L^*$ and brighter galaxies). As mentioned in § 4.1, we expect the AGN clustering amplitude to match that of the entire galaxy population when the mean masses of the two populations are similar.

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

4.3. *Clustering and AGN Activity*

In Figure 3, we show that the 2PCF for the lowest $L_{[\rm O\ III]}$ AGNs in our sample appears to have a higher clustering amplitude than the highest $L_{[\rm O\ III]}$ AGNs. We test whether or not this amplitude difference is a result of the differing AGN host galaxy distributions. From the full AGN sample, we randomly construct two subsamples that possess the same D4000 distributions as shown in Figure 4 for the low and high $L_{[\rm O\ III]}$ samples, but with no regard to the $L_{[\rm O\ III]}$. We find no difference in their respective 2PCFs. We repeat the test for the $u - r$ 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 $L_{[\rm O\ III]}$ samples is driven by the [O III] 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 D4000 distributions, $u - r$ color, e-class, and concentration indices, regardless of $L_{[\rm O\ III]}$. In most cases, we do see differences in the 2PCFs of these subsamples; e.g., the high D4000 sample is more strongly clustered than the low D4000 AGN sample. Likewise, the redder AGNs are more clustered than the bluer AGNs. However, the difference in clustering amplitude is strongest when the AGNs are split by $L_{[\rm O\ III]}$.

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

We thank Volker Springel and Lars Hernquist for their hydrodynamic simulations performed at the Center for Parallel Astrophysical Computing at the Harvard-Smithsonian Center for Astrophysics, Jon Loveday, David Weinberg, Michael Vogeley, and the anonymous referee for useful discussions and/or comments. Funding for the creation and distribution of the SDSS[7] Archive has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the US Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society.

[7] 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, MNRAS, 348, 1355
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, J., Charlot, S., White, S. D. M., Tremonti, C., Kauffmann, G., Heckman, T., & Brinkmann, J. 2004, MNRAS, in press (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, in IAU Symp. 179, New Horizons from Multi-Wavelength Sky Surveys, ed. B. J. McLean, D. A. Golombek, J. J. E. Hayes, & H. E. Payne (Dordrecht: Kluwer), 376
Croft, R. A. C., Dalton, G. B., & Efstathiou, G. 1999, MNRAS, 305, 547
Croom, S., et al. 2003, preprint (astro-ph/0310533)
Davis, M., & Geller, M. J. 1976, ApJ, 208, 13
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
Di Matteo, T., Fabian, A. C., Rees, M. J., Carilli, C. L., & Ivison, R. J. 1999, MNRAS, 305, 492
Dressler, A. 1980, ApJ, 236, 351
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, ApJ, 606, L41
Hamilton, A. J. S. 1988, ApJ, 331, L59
Hao, L., & Strauss, M. A. 2004, in Coevolution of Black Holes and Galaxies, ed. L. C. Ho (Pasadena: Carnegie Obs.), 23
Ho, L. C. 2004, in Coevolution of Black Holes and Galaxies, ed. L. C. Ho (Pasadena: Carnegie Obs.), 293
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
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, Statistics in Theory and Practice (Princeton: Princeton Univ. Press)
Mathur, S. 2000, MNRAS, 314, L17
Mathur, S., Kuraszkiewicz, J., & Czerny, B. 2001, NewA, 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
Oemler, A. J. 1974, ApJ, 194, 1
Peebles, P. J. E. 1980, The Large-Scale Structure of the Universe (Princeton: Princeton Univ. Press)
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
———. 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
———. 2004, ApJ, 608, 16