The merger fraction of active and inactive galaxies in the local Universe through an improved non-parametric classification

Stefano Cotini,[1,2]* Emanuele Ripamonti,[2] Alessandro Caccianiga,[1] Monica Colpi,[2] Roberto Della Ceca,[1] Michela Mapelli,[3] Paola Severgnini[1] and Alberto Segreto[4]

[1]*INAF – Osservatorio Astronomico di Brera, Via Brera 28, I-20121 Milano, Italy*
[2]*Dipartimento di Fisica 'G. Occhialini', Università degli Studi di Milano – Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*
[3]*INAF – Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122 Padova, Italy*
[4]*INAF – Istituto di Astrofisica Spaziale e Fisica Cosmica, Via U. La Malfa 153, I-90146 Palermo, Italy*

ABSTRACT

We investigate the possible link between mergers and the enhanced activity of supermassive black holes (SMBHs) at the centre of galaxies, by comparing the merger fraction of a local sample ($0.003 \leq z < 0.03$) of active galaxies – 59 active galactic nuclei host galaxies selected from the All-Sky *Swift* Burst Alert Telescope (BAT) Survey – with an appropriate control sample (247 sources extracted from the HyperLeda catalogue) that has the same redshift distribution as the BAT sample. We detect the interacting systems in the two samples on the basis of non-parametric structural indexes of concentration (C), asymmetry (A), clumpiness (S), Gini coefficient (G) and second-order momentum of light (M_{20}). In particular, we propose a new morphological criterion, based on a combination of all these indexes, that improves the identification of interacting systems. We also present a new software – pyCASSO (PYTHON *CAS* software) – for the automatic computation of the structural indexes. After correcting for the completeness and reliability of the method, we find that the fraction of interacting galaxies among the active population (20^{+7}_{-5} per cent) exceeds the merger fraction of the control sample ($4^{+1.7}_{-1.2}$ per cent). Choosing a mass-matched control sample leads to equivalent results, although with slightly lower statistical significance. Our findings support the scenario in which mergers trigger the nuclear activity of SMBHs.

Key words: galaxies: active – galaxies: interactions.

1 INTRODUCTION

Observations indicate that the growth history of supermassive black holes (SMBHs; $M_{BH} > 10^6$ M$_\odot$) is closely connected to that of their host galaxies. The discovery of scaling relations, linking the black hole mass to properties of the host in the local Universe, hints for a scenario of galaxy–SMBH symbiotic evolution (Magorrian et al. 1998; Ferrarese & Merritt 2000; Gebhardt et al. 2000; Marconi & Hunt 2003; Häring & Rix 2004; Ferrarese & Ford 2005; Graham 2012a,b). In particular, the near ubiquity of SMBHs in massive spheroids indicates that black hole growth, mainly driven by gas accretion (e.g. Marconi et al. 2004; Croton et al. 2006; Merloni & Heinz 2012; Volonteri & Bellovary 2012), is favoured in galaxies where the importance of organized rotation both in the gaseous and stellar component is weak. As morphological properties of galaxies are likely to be determined by their complex assembly history and can be transient features, the processes that determine the formation and evolution of galaxies affect hand in hand the formation and evolution of SMBHs, and in particular their fuelling.

Theoretical models indicate that galaxy formation and evolution is driven by accretion of gas from the cosmic environment (e.g. Bournaud et al. 2005; Keres et al. 2005; Mapelli, Moore & Bland-Hawthorn 2008; see Sancisi et al. 2008 for a review) and by halo–halo interactions both involving multiple minor mergers or major galaxy–galaxy mergers (e.g. White 1978; Miller & Smith 1980; Gerhard 1981; Negroponte & White 1983; Lake & Dressler 1986; Barnes 1988; Dekel et al. 2009; Bournaud et al. 2011; see Mirabel 2001 for a review). More recently, the mode of gas accretion has been recognized as playing a potentially critical role in shaping galaxies (Sales et al. 2012), leaving open the possibility that spheroids form via multiple episodes of misaligned gas inflows, besides major mergers. In lack of a broad consensus, observations of active galactic nuclei (AGN) and of their galaxy hosts, from suitably selected samples, can provide clues on the mechanisms triggering the SMBH activity, and on their co-evolution.

A longstanding issue is how the gas can lose enough angular momentum from the large scale (~ 0.1–100 kpc) down to the SMBH's

*E-mail: stefano.cotini@gmail.com

horizon scale ($\sim 10^{-5}$ pc). A possible scenario involves gravitational perturbations due to tidal interactions between galaxies in close fly-bys (on ~ 10–70 kpc scales) or/and violent galaxy mergers occurring on smaller scales of $\sim$ kpc or less. These perturbations may drive large quantities of gas towards the centre of the merger remnant (e.g. Kauffmann & Haehnelt 2000; Springel, Di Matteo & Hernquist 2005; Hopkins et al. 2006). This accumulated gas may induce both an intense starburst phase and an enhanced nuclear activity (active SMBH), whose feedback, in turn, can act as a mechanism to regulate subsequent star formation and accretion (Churazov et al. 2001; Best et al. 2006; Schawinski et al. 2006, 2007; McNamara & Nulsen 2007). Galaxy interactions/mergers should be therefore responsible not only for large-scale ($\gtrsim 10^3$ pc) morphological distortions but also for the inflow of gas down to the typical scale of SMBH accretion ($\lesssim 10^{-4}$ pc).

If SMBH activity is triggered, at least partially, by galaxy mergers, the fraction of galaxies with clear sign of being the results of interactions/mergers should be statistically higher in a sample of AGN-host galaxies than in a sample of field galaxies. This and other similar observational tests have been carried out in the last few years with somehow contrasting results (see e.g. Petrosian 1982; Dahari 1984, 1985; Keel et al. 1985; Fuentes-Williams & Stocke 1988; Virani, De Robertis & VanDalfsen 2000; Schmitt 2001; Miller et al. 2003; Grogin et al. 2005; Waskett et al. 2005; Koulouridis et al. 2006; Serber et al. 2006). In particular, while some studies claim a connection between nuclear activity and the presence of close companions or tidal distortions (e.g. Dahari 1984; Keel et al. 1985; Rafanelli, Violato & Baruffolo 1995; Koss et al. 2010, 2011, 2012; Ellison et al. 2011; Ramos Almeida et al. 2011; Silverman et al. 2011; Liu, Shen & Strauss 2012), other studies indicate that there is statistically little support to a AGN-merger connection (Barton, Geller & Kenyon 2000; Schmitt 2001; Dunlop et al. 2003; Grogin et al. 2005; Coldwell & Lambas 2006; Alonso et al. 2007; Ellison et al. 2008; Li et al. 2008; Gabor et al. 2009; Darg et al. 2010; Cisternas et al. 2011; Kocevski et al. 2012).

The differences between various studies might be due to biases in the choice of the galaxy sample. For example, obscured AGN can be missed in studies based on optical emission-line ratios, optical spectral classification or even soft X-ray fluxes. Among the aforementioned studies, only Koss et al. (2010) use a sample of hard X-ray selected AGN, and find a strong excess of merging systems with respect to a control sample.

Another source of error is counting chance superposition galaxy pairs as physically interacting galaxies (for more details about this source of error we refer the reader to section 6.1 of Ellison et al. 2011).

The third source of bias is the possible time delay between the merger and the switch on of the nuclear activity. Various studies (e.g. Ellison et al. 2008; Schawinski et al. 2009, 2010, and references therein) find empirical evidences that mergers enhance star formation first, and only at later epochs trigger the AGN phase (~ 500 Myr after the starburst). In fact Smirnova, Moiseev & Afanasiev (2010) analyse a sample of apparently isolated Seyfert galaxies and find that about 35 per cent of them show tidal tails, consistent with a gas-rich merger (likely a minor merger) in the last 0.5–1 Gyr. Thus, samples of galaxy pairs might miss, by default, late merger phases and gas-rich minor mergers. This problem is less acute when empirical measures of galaxy morphology are used, as they can identify a galaxy as the result of an interaction/merger even when it lacks a companion (provided that interaction features are strong enough). Therefore, these measures are sensitive both to the initial and the late stages of mergers, and are less biased against specific merger phases.

In this paper we re-address the possible link between mergers/interactions (in the following, we will use the two terms as synonymous) and SMBH activity, by comparing the merger fraction of an AGN host galaxy sample to the typical merger fraction of galaxies in the local Universe.

To satisfy the need that both the galaxy sample and the method of analysis are as unbiased as possible, (i) we use a hard (>10 keV) X-ray selected AGN sample (not to miss obscured sources, with the partial exception of the heavily absorbed Compton thick AGN, i.e. those sources with absorbing column densities exceeding 10^{24} cm^{-2}), and (ii) we adopt a non-parametric morphological analysis (to identify truly interacting galaxies even in late merger phases).

Moreover, we propose an improved technique for evaluating the merger fraction of a galaxy sample by using a method that is objective, reliable and fast, so that it can be applied, in the future, to larger samples of galaxies; we also define the completeness and the reliability coefficients, that allow a statistical correction of the merger fraction and further reduce possible residual errors in the automated classification.

This paper is organized as follows: Section 2 presents the galaxy samples and the procedure adopted for their unbiased selection; Section 3 explains the non-parametric morphological method used for the analysis; Section 4 presents our estimates of the merger fraction of the AGN Burst Alert Telescope (BAT) sample and of the control sample; Section 5 outlines a summary of the most important points. Appendices A and B present, respectively, the data processing algorithms (including a detailed description of the software that we developed for our automated classification) and a discussion on the image degradation effects that affect data analysis.

2 SAMPLE SELECTION

The aim of this work is to study the possible link between mergers and SMBH activity, by comparing the merger fraction of an AGN host galaxy sample to the typical merger fraction of galaxies in the local Universe. To this purpose, we select two samples: the first one is a hard (15–195 keV) X-ray selected sample of active galaxies (will be addressed here as the BAT sample), which is similar – with several objects in common – to the sample already used in Koss et al. (2010). The second one is an optically selected control sample of galaxies (without any imposition on their active nature) that we extract from the HyperLeda catalogue (Paturel et al. 2003). We impose on both samples a minimum redshift of 0.003 to avoid too extended sources (image processing faces some difficulties in these cases) and a maximum redshift of 0.03, because the optical counterparts of the selected galaxies need to match the requirements for our morphological analysis (see Appendix A3 in the electronic edition).

2.1 BAT sample

The BAT is a coded aperture imaging camera on-board the *Swift* satellite (Gehrels et al. 2004); it has a wide field of view (1.4 sr), a point spread function (PSF) of 17 arcmin [full width at half-maximum (FWHM)] and it operates in the 15–195 keV energy range. To select a sample of AGN out of *Swift* BAT observations, we adopt the Palermo *Swift*-BAT hard X-ray catalogue (Cusumano et al. 2010) that collects the data relative to the first 54 months of the *Swift* mission and is therefore one of the most complete, well defined and extended catalogues of hard X-ray sources up to date. It contains 1256 sources with a signal-to-noise ratio greater than 4.8, a flux limit

of 6.0×10^{-12} erg cm^{-2} s^{-1} and a counterpart identification with a 95 per cent confidence level. This catalogue represents a relatively unbiased sample of AGN, because it is based on a particular hard X-ray band, where biases against absorbed AGN are less important. For our analysis, we extract from this catalogue a complete sample of 523 sources, with absolute Galactic latitude $|b| > 15°$, S/N > 5 and flux greater than 8.0×10^{-12} erg cm^{-2} s^{-1}. Secondly, we select a complete subsample in the redshift interval $0.003 \leq z < 0.03$ and, finally, we restrict to the area of sky covered by the Sloan Digital Sky Survey Data Release 8 (SDSS DR8, http://www.sdss3.org/dr8/; Aihara et al. 2011), to make use of the optical data offered by this survey. The final BAT active galaxy sample[1] consists of 59 sources (15 at redshift $0.003 \leq z < 0.01$, 16 at redshift $0.01 \leq z < 0.02$ and 28 at redshift $0.02 \leq z < 0.03$), which represent $\sim$35 per cent of the total number of galaxies belonging to the complete sample in the same redshift interval 0.003–0.03 (169 objects).

The BAT sample is not a mere selection of galaxies, but of systems instead: the sources are selected on the basis of the presence of one AGN at least, but the poor angular resolution of *Swift* BAT observations does not allow us to distinguish the possible X-ray emission of multiple AGN in pairs or group of galaxies. As a consequence, in the case of merging galaxies, the ensemble of objects is considered as a single (interacting) system, likewise each isolated galaxy represents a single (but non-interacting) system. In particular, the 'interacting' or 'non-interacting' classification is determined from the results of the automated structural analysis (see Section 3.2).

2.2 The control sample

The control sample is used to evaluate the average merger fraction among galaxies and to compare it with the same value found in the BAT sample (i.e. among AGN), so it has to match the redshift distribution of the BAT sample and it must be unbiased towards interacting or isolated systems.

For example, a random sampling among SDSS galaxies would lead to an overestimate of the merger fraction, because interacting systems have more chances to be selected than isolated galaxies (in fact they can be sorted out by each one of their members). Therefore, we replicate the particular 'system classification' of the BAT sample also in the control one. In the following we describe the procedure used to define the control sample.

(i) We select three random square boxes of sky fully covered by SDSS imaging. All boxes have a side of $7°.5$ and contain, on average, $\sim$300 galaxies[2] in the $0.003 \leq z < 0.03$ redshift interval. The choice of multiple medium-size boxes, instead of a single large box, avoids biases related to local peculiar environment (i.e. galaxy groups or clusters). The size of the boxes ensures a significant number of sources inside each one, so that possible border effects become unimportant (i.e. the loss of one galaxy of a pair that lies halfway the edge of the box).

(ii) We consider all the sources in the HyperLeda catalogue present in the three boxes of sky quoted above. For each galaxy, we acquire the SDSS image and, on the basis of the structural parameters (asymmetry, clumpiness, Gini coefficient, second-order

momentum of light – see Section 3), we distinguish whether it is interacting or isolated.

(iii) We switch from the 'galaxy classification' to the 'system classification': we consider as a multiple system every ensemble[3] of sources in which at least one galaxy has been identified as interacting through our classification. We consider only one galaxy of each system, so that each system is represented only by one entry. At this point, the control sample consists of 734 systems (79 at redshift $0.003 \leq z < 0.01$, 67 at redshift $0.01 \leq z < 0.02$ and 588 at redshift $0.02 \leq z < 0.03$).

(iv) The redshift distribution of these sources is considerably different from the BAT sample, so possible redshift-related effects (i.e. an evolution of the merger fraction) may alter significantly our comparison. For this reason, we reduce our sample by randomly extracting, in each redshift bin, the right number of sources to match the BAT sample's redshift distribution.

At the end of this procedure we obtain a redshift-matched control sample of 247 sources, distributed as shown in Table 2, that are fully comparable with the BAT ones. We point out that the control sample contains both active and quiescent galaxies at random, because we want to check whether the merger fraction of the BAT AGN sample is significantly higher than the typical merger fraction in the local Universe.

3 DATA ANALYSIS

In this work we aim to determine the merger fraction of two samples of galaxies using a method that is objective, reliable and fast, so that it can be applied, in the future, to larger samples of galaxies. To this end, different techniques have been developed in the last decades (i.e. Byun & Freeman 1995; Abraham et al. 1996; Le Févre et al. 2000; Patton et al. 2000, 2002; Peng et al. 2002; Blanton et al. 2003; Conselice 2003; Lotz, Primack & Madau 2004; Scarlata et al. 2007) but there is not yet a method that has been proven to be clearly superior to the others.

Pair counts require a strong observational effort, because they need redshift measurements for each galaxy, to avoid chance superpositions. Moreover, even pairs of galaxies at the same redshift could be not gravitationally bound, leading to an overestimate of the merger fraction.

Other techniques rely on the identification of galaxies that, due to gravitational interactions with a close companion, show morphological perturbations. The visual, qualitative, classification is the most used and accurate method, but it is intrinsically subjective and becomes less and less reliable with increasing redshift, because of the lower resolution and S/N ratio. Moreover, it is time consuming, and, therefore, it is not applicable on very large samples of galaxies.

Quantitative classifications are less accurate but more objective, and allow corrections for high redshifts, because the image degradation is measurable. Among these, we can distinguish between parametric and non-parametric classifications. In the first kind, the projected light distribution of the galaxy is either fitted as a whole with an analytical model (like the Sérsic or the de Vaucouleurs profile), or it is split in its various components (i.e. a bulge and a disc)

---

[1] This sample does not include two sources that have too low resolution for being analysed and one source that is very close to a bright star, which invalidates our analysis.

[2] We point out that, due to our subselections and the impossibility to analyse all the images, the number of valid objects in each box is usually reduced almost by 20 per cent.

[3] The extent of the ensemble depends on the number and the kind of sources falling into the aperture (automatically computed on the basis of the light profile of the central galaxy, see Appendix A3 in the electronic edition) for the estimation of structural indexes. However, in general, it is unlikely that galaxies with separation greater than 30 kpc are included in the same aperture.

that are fitted separately. Nevertheless, these methods are quite unsuitable for irregular or disturbed galaxies and, in the case of close pairs, the subtraction of the extra light coming from the companion is not trivial. Non-parametric techniques are not based on any analytical models, so they are equally applicable on every kind of galaxy; however, it is more difficult to convert their values into physically meaningful results. An interesting, non-parametric classification has been developed in recent years by Conselice (2003) and Lotz et al. (2004): it consists in a set of five structural indexes that measure specific properties of a galaxy. The first three parameters, concentration (C), asymmetry (A) and clumpiness (S), presented by Conselice (2003), are referred as the *CAS* system; the other two indexes, introduced by Lotz et al. (2004), are the Gini coefficient (G) and the second-order momentum of light (M_{20}). We decide to adopt this non-parametric approach for our analysis and we will refer to the whole set of indexes as *CASGM* system. As in a visual analysis, the *CASGM* method becomes less reliable in the case of low resolution or S/N ratio, but these effects have been well quantified by Lotz et al. (2004) and are reported at the end of Appendix A3 (in the electronic edition). Taking into accounts this limits, we have imposed a maximum redshift of 0.03 to our samples, so that SDSS images ensure the minimum requirements for the automated analysis.

3.1 The *CASGM* parameters

In order to compute these parameters, we need first to determine the extension of the galaxy, which is based on the Petrosian radius. The Petrosian index of a galaxy is the ratio between the mean surface brightness inside radius R, $\bar{\mu}(r < R)$, and the surface brightness $\mu(R)$ at R, that is

$$\eta(R) = \frac{\bar{\mu}(r < R)}{\mu(R)}. \tag{1}$$

The Petrosian radius is the radius r_P at which the inverted Petrosian index is equal to 0.2 (Petrosian 1976). For the *CAS* system, the area of the galaxy is the circular area inside 1.5 times the Petrosian radius at $r(\eta = 0.2)$, with centre in the point that minimizes the asymmetry of the galaxy.

(i) *Concentration.* The concentration index is the ratio of the light inside an inner aperture (circular or elliptical) to the light inside an outer aperture. The *CAS* system adopts the Bershady, Jangren & Conselice (2000) definition, so C is defined as

$$C = 5 \log \left(\frac{r_{80}}{r_{20}} \right), \tag{2}$$

where r_{80} and r_{20} are the radii that contain the 80 and the 20 per cent of the total light of the galaxy, respectively. Typical values of C range from ∼2 to ∼5: elliptical galaxies and spheroidal systems usually have $C > 4$, disc galaxies have $3 < C < 4$, while galaxies with a low surface brightness or a low velocity dispersion have $C \sim 2$.

(ii) *Asymmetry.* The A coefficient measures the asymmetry degree of the galaxy light distribution under a 180° rotation. This index was originally used to describe galaxy morphologies (i.e. Abraham et al. 1996), but we follow the slightly modified formulation of Conselice, Bershady & Gallagher (2000). This index is computed by subtracting the 180° rotated image to the original one, and by normalizing the residuals by the total flux of the galaxy. This value is then corrected by subtracting the asymmetry contribution of the background (i.e. produced by a luminosity gradient or a close stellar

halo), which is computed in the same way. Therefore, the final value of A is

$$A = \frac{\sum_{i,j} |I(i, j) - I_{180}(i, j)|}{\sum_{i,j} |I(i, j)|} - \frac{\sum_{i,j} |B(i, j) - B_{180}(i, j)|}{\sum_{i,j} |I(i, j)|}, \tag{3}$$

where I and B are, respectively, the original image of the galaxy and of the background, while I_{180} and B_{180} are their rotated images. This coefficient is sensitive to all the processes that introduce a certain degree of asymmetry in the light distribution, such as star-forming regions, dust bands and mergers. The relative contribution of these elements has been studied by Conselice (2003), who showed that small-scale structures can make up only to the 30 per cent of the asymmetry of the galaxy; therefore, A is dominated by large-scale effects and is a good tracer of mergers and gravitational distortions.

(iii) *Clumpiness.* The S index has been introduced by Conselice (2003) to quantify the patchiness of the galaxy, that is the fraction of light coming from small-scale structures, such as clumps of star formation. It is defined as the ratio of the flux contained in high-frequency features to the total flux of the galaxy. It is computed by subtracting a blurred[4] copy of the image to the original one, and then normalizing by the total flux of the galaxy. The value is then corrected by removing the background clumpiness, so it is equal to

$$S = 10 \frac{\sum_{i,j} (I(i, j) - I_S(i, j))}{\sum_{i,j} I(i, j)} - \frac{\sum_{i,j} (B(i, j) - B_S(i, j))}{\sum_{i,j} I(i, j)}, \tag{4}$$

where I and B are the original image of the galaxy and of the background, respectively, while I_S and B_S are their blurred images. The nuclear $0.25r_P$ region is excluded from the computation, because it would give a high clumpiness contribution, which is not related to a region of young and intense star formation. Moreover, negative values after the subtraction of the smoothed image are forced to zero (Conselice 2003).

Large values of S indicate that most of the light of the galaxy is accumulated in few and clumpy structures (i.e. starburst galaxies), while low values of S indicate that the light distribution is smooth (i.e. elliptical galaxies).

G and M_{20} are based on the segmentation map of the galaxy defined by Lotz et al. (2004). In contrast with the circular and the elliptical apertures of the *CAS* indexes, the segmentation map can assume any irregular shape because its constraints (see Appendix A3.7 in the electronic edition) are only a brightness limit (to exclude the background and possible spurious pixels) and a continuity requirement (any source that is not directly connected with the galaxy is not taken into account). Therefore, the segmentation map can follow accurately the outline of the galaxy, especially in the case of close couples and mergers.

(iv) *Gini coefficient.* The Gini coefficient is a measure of statistical dispersion. It is usually adopted in economics to describe the inequality of a distribution (i.e. levels of income) and was adapted by Abraham, van den Bergh & Nair (2003) and Lotz et al. (2004) for the morphological classification of galaxies. The formulation of the Gini coefficient is based on the Lorentz curve:

$$L(p) = \frac{1}{\bar{X}} \int_0^p F^{-1}(u) \, du, \tag{5}$$

---

[4] The blurring is obtained by convolving the original image with a filter of width $\sigma = 0.25r_P$.

where p is the percentage of the faintest pixels, $F(x)$ the cumulative distribution function and $\bar{X}$ the average value of all the X_i intensities. After some rearrangements (Glasser 1962) and a correction to compensate for the Poissonian noise in the faintest regions of the galaxy (Lotz et al. 2004), it can be expressed as

$$G = \frac{1}{|\bar{X}|n(n-1)} \sum_i^n (2i - n - 1)|X_i|. \tag{6}$$

The Gini coefficient is computed on the segmentation map and represents a sort of generalized concentration index, in fact it tells whether the light is evenly distributed inside the galaxy, but does not depend on any particular centre. This index can range from zero, in the case of a perfectly uniform distribution, to one, in the case that all the light of the object is concentrated in a single pixel.

(v) *Momentum of light.* The M_{20} coefficient measures how far from centre are located the brightest pixels of the galaxy. It is based on the total second-order momentum of light M_{tot}, that is the sum, over all the pixels of the segmentation map, of the pixels' flux f_i multiplied for its square distance from the centre:

$$M_{tot} = \sum_i^n M_i = \sum_i^n f_i \left[(x_i - x_c)^2 + (y_i - y_c)^2 \right]. \tag{7}$$

The x_c and y_c variables are the coordinate of the galaxy centre, which is now defined as the pixel that minimizes the value of M_{tot}. The M_{20} coefficient is the second-order momentum of the brightest 20 per cent of the galaxy. To compute it, we follow the procedure in Lotz et al. (2004): the pixels of the segmentation map are sorted by decreasing flux; then the corresponding momenta M_i are summed, until the sum of the brightest pixels equals 20 per cent of the total galaxy flux; finally this value is normalized by M_{tot}, so

$$M_{20} = \log_{10} \left(\frac{\sum_i M_i}{M_{tot}} \right) \quad \text{while} \quad \sum_i f_i < 0.2 f_{tot}. \tag{8}$$

The normalization removes dependencies on the size of the object and its total flux, making M_{20} less subject to inclination effects. Being weighted on the square of the distance from the centre, this index is especially suitable for detecting double nuclei systems (such as close galaxies in a merging phase), because the brightest pixels of the system are off-centre and they give a large contribution to the value of M_{20}.

The *CASGM* system relies on the Petrosian radius, that, being based on a curve of growth, is independent of the galaxy size and largely insensitive to both the S/N ratio and the surface brightness of the sources (see Lotz et al. 2004 for a discussion about the influence of low S/N and resolution on these parameters).

These indexes are related to galaxy morphologies, and the authors of the *CASGM* system have calibrated a complete classification using the Frei et al. (1996) catalogue. Moreover, linking couples of *CASGM* indexes, they defined some fiducial sequences, that allow the separation of normal[5] and merging galaxies. In our work we use the two[6] main merger criteria:

(i) *A–S criterion* (Conselice 2003). In the plane A versus S normal galaxies show a good correlation:

$$A_{fit}(R) = (0.35 \pm 0.03)\, S(R) + (0.02 \pm 0.01). \tag{9}$$

The two indexes are computed on R-band images, because they are less sensible to bright young stars and provide a more stable relation. Mergers should deviate from this relation because their light distribution, distorted from gravitational interactions, raises significantly the value of A, while it has a weaker influence on the S parameter. Therefore, galaxies that show a large deviation from the fiducial sequence, or simply a very high value of asymmetry, that is

$$A > A_{fit} + 3\sigma \quad \text{or} \quad A > 0.35, \tag{10}$$

are classified as mergers (σ is the mean dispersion in equation 9, and is equal to 0.035).

(ii) *G–M_{20} criterion* (Lotz et al. 2004). As in the previous case, the correlation among normal galaxies in the plane G versus M_{20} is used to define this merging criterion:

$$G > -0.115\, M_{20} + 0.384. \tag{11}$$

3.2 Data processing

Our data processing workflow is organized into three main steps, each one coupled with a specific software.

(1) *Data acquisition.* SDSS frames cover a field of view of $\sim 14 \times 10$ arcmin^2. Because our galaxies are near and extended, they are often close to the edge of the image, or they fall halfway along multiple frames. We use the software MONTAGE[7] to assemble multiple images in FITS format (details about this step are given in Appendix A1 in the electronic edition).

The HyperLeda data base is an ideal starting catalogue for this operation since it provides, for each galaxy, the list of properties (coordinates, diameter, position angle, redshift etc.) to automatically run MONTAGE. Because we are still dealing with a moderate number of sources, we carefully checked the correct assembly of all the images.

(ii) *Pre-processing.* In this step we prepare the image for the computation of the structural indexes: every feature that might affect the *CASGM* analysis (i.e. bright stars in foreground, cosmic rays, image artefacts etc.) must be masked. For our automated workflow, we used the software SEXTRACTOR (Bertin & Arnouts 1996) that provides a fast detection of all the sources in the image. Source identification is essentially based on local intensity and contrast, but the software examines also the light profile, extracting a number of properties (for a detailed description see Appendix A2 in the electronic edition). Therefore, at the end of the pre-processing step,[8] the original image is associated with a SEXTRACTOR catalogue, and to several 'service' images, that specify the regions to exclude and provide useful information for the *CASGM* analysis.

(iii) *CASGM analysis.* The crucial part of this work is entrusted to our software PYCASSO (PYTHON *CAS* Software), whose task is to

---

[5] Throughout the rest of this text, we will address non-merging galaxies as normal.

[6] We tried also the relation based on the asymmetry and the Gini coefficient (Lotz et al. 2004), but it gave a worse subdivision of the merging systems. Therefore, we rejected this relation.

[7] Developed by the NASA Earth Science Technology Office; http://montage.ipac.caltech.edu/

[8] After the pre-processing phase, about 12 per cent of the sources is discarded, usually because the HyperLeda coordinates are wrong, or MONTAGE cannot produce the mosaic or it is impossible to set-up the image properly (i.e. because the galaxy is too faint and is not fitted correctly by SEXTRACTOR).

compute the *CASGM* structural indexes. PYCASSO is entirely developed in PYTHON, an high level and object oriented programming language, with extensive standard library and the possibility to import modules[9] suited for handling scientific data and astronomical images. We give a detailed description of the algorithms implemented in PYCASSO in Appendix A3 (in the electronic edition).

We tested our workflow and softwares on the Frei et al. (1996) catalogue. This catalogue collects a sample of nearby, well-resolved galaxies, and it is therefore suitable for testing the reliability of the algorithm, possible side effects (see Appendix B in the electronic edition for an image degradation discussion) and improvements in the implementation of the *CASGM* indexes. We compared our results on these galaxies with Conselice (2003), Lotz et al. (2004) and Vikram et al. (2010) and we found a very good agreement: on average, the *C*, *A*, *S* and M_{20} coefficients are consistent within 1σ with the results of the other authors, while the Gini coefficient is in agreement within 1.5σ.

To further test the *CASGM* method, we carried out a visual classification on all the systems identified as merger by the *CASGM* analysis, both in the BAT and in the control samples. The visual classification assigns each galaxy to one of these three classes: (i) 'normal' galaxies do not show any signs of interaction (i.e. appear regular and isolated); (ii) 'edge-on' galaxies: these are intentionally kept separated from non-edge-on galaxies to study possible biases related to dust bands, as highlighted by other studies (i.e. De Propris et al. 2007; Jogee et al. 2009); (iii) 'merger' systems, i.e. close pairs of galaxies and sources showing morphological distortions or perturbations (such as tidal tails, double nuclei etc.). The visual classification is based first on the *RGB* and FITS images available in the SDSS data base and on the corresponding FITS images. Where available, we exploited also the spectroscopic data to discern projected pairs of galaxies from real ones. Finally, for the most critical objects, we searched for further information in NED[10] (Gesch et al. 2002; Gesch 2007) and in the literature.

4 RESULTS

Here we present the results of our automated classification and the merger fraction of the two samples. As explained in Section 2.2, the BAT sample is a collection of systems, so we have to switch from galaxy to system classification also in the control sample, to make them fully comparable. To this purpose we consider as a single system any ensemble of galaxies for which one galaxy, at least, has been classified as interacting. The interacting or non-interacting classification is of course provided by the specific merger criterion considered.

We ran PYCASSO using both elliptical and circular apertures and we visually checked the control images and the results produced by our software. In most of the cases the two analyses coincide, but for some class of objects (i.e. edge-on galaxies and mergers) the elliptical apertures prove to be more reliable, being able to better fit the outline of these sources. In the case of stretched objects, instead, circular apertures include a large amount of background, so the corrections applied to the asymmetry and the clumpiness become more critical. For these reasons, we report only the results of the elliptical classification.

---

[9] In particular, PYCASSO needs NUMPY scientific module, essential for matrix operations; PYFITS used to read images in FITS format and MATPLOTLIB used to create control images and plots.
[10] NASA/IPAC Extragalactic Database, http://ned.ipac.caltech.edu/

Lotz et al. (2004) studied the typical errors associated with the *CASGM* measurements by analysing the Frei et al. (1996) images and the SDSS images of the same galaxy sample. These differences provide an average estimate of the uncertainties on the indices, in fact (i) they take implicitly into account the slight smoothing effect introduced by MONTAGE (because the Frei et al. 1996 galaxies always belong to a single frame); (ii) they take into account the differences due to image resolution and quality (because Frei et al. 1996 have a lower resolution, so they are similar to SDSS images at larger redshifts). The uncertainties related to the structural indexes are the following: $\delta C = 0.11$, $\delta A = 0.04$, $\delta S = 0.09$, $\delta G = 0.02$ and $\delta M_{20} = 0.12$.

4.1 Results of the BAT sample

The results obtained on the BAT sample, using both the visual and the automated classifications, are reported in Table 1 and discussed in the following (errors on the merger fractions are of 68 per cent confidence level and have been computed using the Gehrels 1986 prescriptions).

Table 1. Merger identifications in the BAT sample: we report only those galaxies that have been tagged as interacting by at least one classification method. In the visual classification, '*M*' identifies mergers, '*n*' the non-interacting galaxies and '*e*' the edge-on galaxies. The mergers of the automated criteria are labelled by a '×' mark. It is possible to notice that the combined criterion is much more reliable than the others, in fact it removes most of the contaminations and it provides results in good agreement with our visual analysis.

| PBCJ name	Visual analysis	Automated classification		
		A–S	G–M_{20}	Combined crit.
0042.8−2331	*M*	×	×	×
0124.4+3346	*M*			
0209.4−1010	*M*	×	×	×
0241.5+0709	*M*	×		
0252.4−0832	*e*		×	
0255.2−0011	*M*	×	×	×
0303.8−0107	*n*	×		
0742.4+4498	*n*	×		
0744.1+2915	*M*	×	×	×
0759.9+2324	*n*			
0823.0−0454	*M*	×	×	×
0919.9+3712	*e*		×	
0926.1+1245	*n*	×		
0942.1+2342	*n*	×		
1002.0+5539	*e*	×	×	×
1023.5+1951	*M*	×	×	×
1104.4+3813	*M*	×	×	×
1113.7+0930	*n*	×		
1139.6+3157	*M*	×	×	×
1204.4+2018	*n*	×		
1206.2+5244	*n*	×	×	×
1217.1+0712	*e*		×	
1225.7+1240	*e*		×	
1345.4+4141	*e*		×	
1417.9+2508	*n*	×		
1424.3+2436	*n*		×	
2236.0+3358	*M*	×	×	×
2318.9+0014	*M*	×	×	×
Total mergers	12	20	18	12
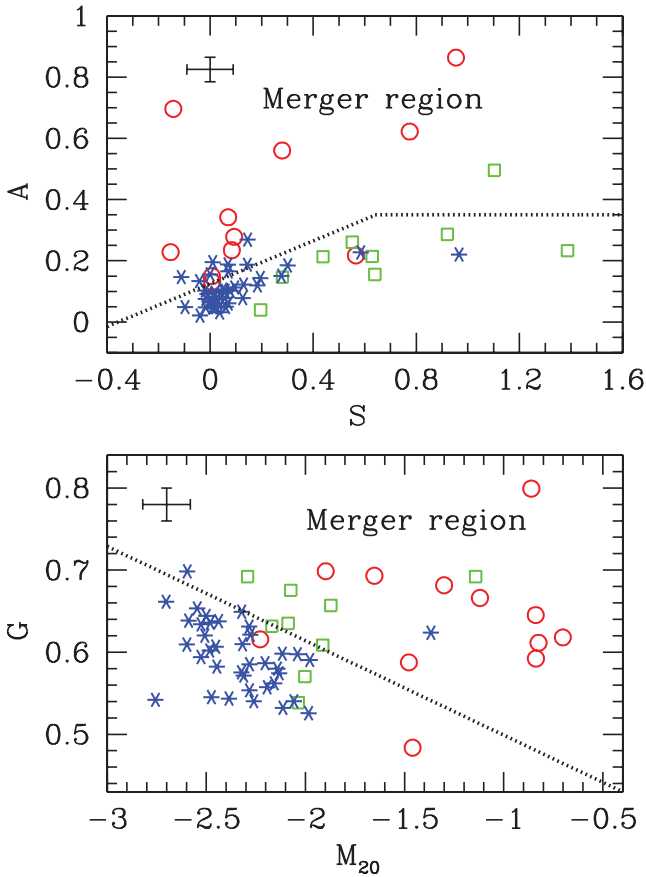
Figure 1. Comparison between visual and automated classifications. Red circles: galaxies visually classified as interacting; green squares: edge-on galaxies; blue asterisks: normal galaxies. Structural indexes classify as merger the galaxies lying above the dotted lines. The error bars are average differences between SDSS and Frei observations of the same objects (Lotz et al. 2004). The A–S criterion shows a slight contamination produced by normal galaxies, while the G–M_{20} is biased towards edge-on galaxies.

(i) *Visual classification.* Through the visual classification we estimate a merger fraction of 20^{+7}_{-5} per cent (we identify 12 mergers, nine edge-on galaxies and 38 normal systems).

(ii) *A–S classification.* The criterion based on the asymmetry and the clumpiness (equation 10) detects 20 mergers, giving a merger fraction of 34 ± 7 per cent. 11 of the 12 systems visually classified as merger have the same classification with the A–S method (see Table 1 and Fig. 1, upper panel). The higher fraction of mergers detected with the A–S method is due to a moderate contamination of normal systems with low clumpiness. In fact, for these cases, even a small asymmetry contribution, produced by small spurious[11] sources within the *CAS* aperture, may be enough for labelling that galaxy as interacting.

(iii) *G–M_{20} classification.* The criterion based on the Gini coefficient and the momentum of light (equation 11) identifies 18 mergers, giving a merger fraction of 31 ± 7 per cent. In this case, the higher fraction of merging systems with respect to the visual classi-

---

[11] For example, in some cases SExtractor is not able to separate the faint high-redshift galaxies in the background from the main one, and the same occurs for small stars in the foreground. If the clumpiness value is near zero, the asymmetry contribution coming from these sources may determine their misclassification as mergers.
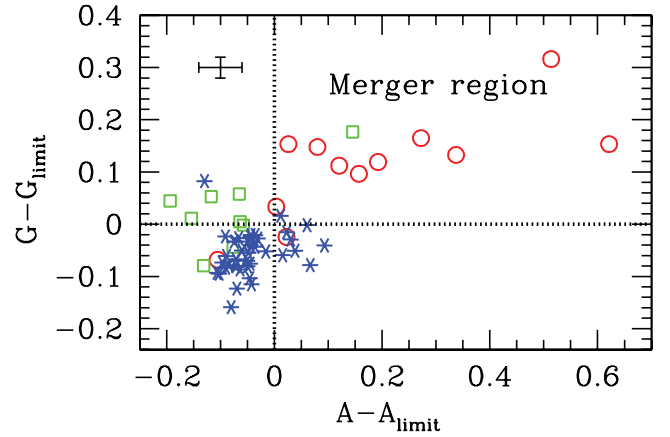

Figure 2. Comparison between our combined criterion and the visual classification: the structural indexes classify as merger the galaxies lying in the top right-hand sector, while symbols and colours are the same as in Fig. 1. The combined criterion shows a good agreement with our classification, in fact the contaminations affecting the original criteria are almost completely removed.

fication is due to the contamination produced by edge-on galaxies. These galaxies are observed through dust bands that obscure the central part of the source and leave two bright areas symmetrically off-centred that influence the momentum of light. A similar effect occurs also for pronounced barred galaxies. 10 of the 12 systems visually classified as mergers have the same classification also through the G–M_{20} method (see Table 1 and Fig. 1, lower panel).

4.1.1 Improvement of the CASGM system

As shown in the previous section, the automated classifications correctly identify almost all the interacting systems, but they systematically overestimate the real number of mergers. For this reason, we introduce an advanced criterion that blends together[12] the previous procedures: we consider as mergers only those systems that satisfy simultaneously the A–S and the G–M_{20} criteria. All these indexes have similar resolution and S/N requirements and so they can be used together; however, this choice may limit the effectiveness of the merger identification, because each method is not sensible to the entire duration of the merger and the interaction phases mapped by each criterion do not fully overlap (see Conselice 2006; Lotz et al. 2008). We expect the combined criterion to be much more reliable than the original ones. For instance, the G–M_{20} contamination should be largely removed because edge-on and barred galaxies are basically symmetric and, therefore, they should be excluded by adding the A–S classification.

4.1.2 Merger fraction of the BAT sample

The combined criterion proves to be an optimal solution, in fact it does not miss almost any merger compared to the previous criteria and it removes about 77 per cent of their wrong classifications, leading to a merger identification in excellent agreement with our visual analysis (see Table 1 and Fig. 2). By exploiting the combined

---

[12] Some attempts have been made by Lotz et al. (2004), who studied other combinations of *CAS* and *GM* indexes (i.e. G–A, G–S, A–M_{20}) which, however, did not produce better classifications than the A–S and G–M_{20} criteria.

Table 2. Detailed comparison of the merger fraction f_m of the BAT and of the control sample in each redshift bin, according to the classification of the combined criterion. In the 'Total (*CASGM*)' row we summarized the results of the mere *CASGM* classification, while in the 'Corrected' row we indicate the merger fractions after the application of the reliability and the completeness corrections. AGN host galaxies are found more frequently in phase of interaction compared to a random selection of galaxies in the same redshift interval. This suggests that there is a link between the merging event and the activity of the SMBH at the centre of galaxies.

Redshift	BAT sample			Control sample		
	N_{sys}	N_m	f_m	N_{sys}	N_m	f_m
$0.003 \leq z < 0.01$	15	3	20 (9–36)	63	3	4.8 (2.2–9.2)
$0.01 \leq z < 0.02$	16	3	19 (9–34)	67	4	6.0 (3.1–10.5)
$0.02 \leq z < 0.03$	28	6	21 (13–32)	117	9	7.7 (5.2–11.0)
Total (*CASGM*)	59	12	20 (15–27)	247	16	6.5 (4.9–8.5)
Corrected	59	12	20 (15–27)	247	10	4.0 (2.8–5.7)

criterion, we detect 12 disturbed systems, so the merger fraction of the BAT sample is 20^{+7}_{-5} per cent.

Even if the low statistics does not allow any strong conclusions, we point that the merger fraction among each redshift bin is almost constant (Table 2), so it does not display any evident signs of evolution in the local Universe.

Our results are in excellent agreement with Koss et al. (2010), who performed a visual analysis on a similar BAT subsample and found a merger fraction of 25 per cent, by considering all the perturbed galaxies and the pairs with a separation below 30 kpc. We have compared the luminosity distributions (14–150 keV band) of the interacting and the non-interacting systems of our BAT sample and, according to a Kolmogorov–Smirnov (KS) test ($\text{prob}_{ks} = 0.2$), the luminosity distributions of the two subsamples do not display significant differences.

4.1.3 Statistical corrections

It is possible to further improve our results by estimating the completeness and the reliability of the automated classification and applying a statistical correction to the merger fraction.

(i) *Completeness.* It quantifies the amount of missed mergers, that is the number of systems that have been labelled as 'interacting' by the visual classification, but as 'non-interacting' by the combined criterion. We define this coefficient as

$$C_{CASGM} = \frac{N_{m,true}}{N_{m,visual}}, \qquad (12)$$

where $N_{m,true}$ is the number of mergers in common between the automated and the visual classification, while $N_{m,visual}$ represents the number of mergers of the visual classification. By extrapolating the completeness from the BAT sample, we obtain $C_{CASGM} = 10/12 = 0.8^{+0.1}_{-0.2}$. This parameter allows us to derive the real merger fraction of the sample, in fact it tells that the number of mergers that have been correctly[13] detected by the automated classification is about 80 per cent of the real number.

(ii) *Reliability.* It quantifies the fraction of normal systems that have been erroneously classified as mergers by the automated procedure. We define it through the probability, P, that the procedure
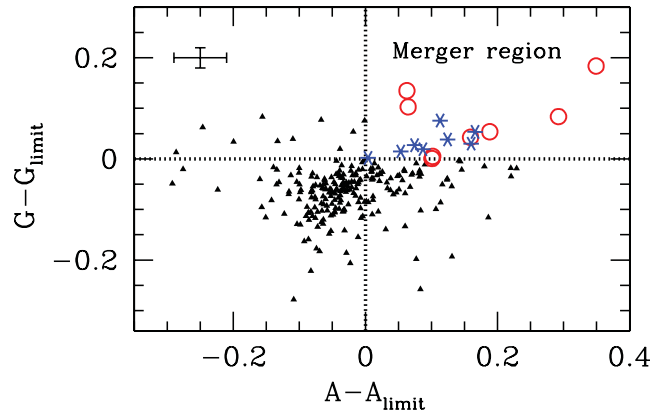
---

[13] Spurious and wrong merger detections must be excluded from the sum.

Figure 3. Classification of control sample's systems according to the combined criterion: the systems lying in the top right-hand sector of the plot are labelled as mergers by the automated criterion. For these objects, we performed a visual analysis (red circles: galaxies visually classified as interacting; blue asterisk: normal galaxies), while black triangles represent non-merger systems according to *CASGM*.

gives a false positive (false merger) in the case of a non-merging system, i.e.

$$P_{CASGM} = \frac{N_{m,false}}{N_{normal}}, \qquad (13)$$

where $N_{m,false}$ is the number of wrong mergers and N_{normal} is the number of non-interacting sources (that is the difference between the number of systems N_{sys} in the sample and the number of real mergers $N_{m,real}$). By extrapolating this value from the BAT sample, we obtain: $P_{CASGM} = 2/47 \sim 0.04^{+0.06}_{-0.03}$, which means that about 4 per cent of the non-interacting systems is instead classified as merger by the combined criterion.

A good knowledge of these coefficients is extremely useful for correcting the merger fraction of very large samples that cannot be visually inspected. In fact, by applying the reliability correction, we obtain the number of 'true' mergers detected by the software, and then, taking into account the completeness coefficient, we can estimate the real number of interacting systems $N_{m,real}$:

$$N_{m,real} = \frac{N_m - P_{CASGM} N_{sys}}{C_{CASGM} - P_{CASGM}}, \qquad (14)$$

where N_m is the number of mergers detected by the combined criterion and N_{sys} is the total number of systems in the sample.

4.2 Results of the control sample

4.2.1 Merging fraction and statistical corrections

The procedure described in the previous sections detects 16 merging systems in the control sample (see Table 2 and Fig. 3) corresponding to a merger fraction of $f_{m,control} = 6.5^{+3.0}_{-1.6}$ per cent. This fraction, however, does not take into account the corrections for the reliability and the completeness previously discussed. Using our estimates of P_{CASGM} and C_{CASGM} based on the BAT sample, we derive that the real number of mergers in the control sample is (see equation 14) $N_{m,real} \sim 8$. In particular, the expected number of true mergers among the 16 detected by the algorithm is ~ 6 (P_{CASGM} correction), while two more real mergers are expected to be missed by the procedure (C_{CASGM} correction). Given the large fraction of the detected mergers that are expected to be spurious (more than 60 per cent),

we have visually inspected all the 16 systems found by the procedure as mergers to confirm and better constrain the actual number of false/true mergers. In good agreement with our expectations, we find that only eight systems are true mergers, the remaining ones being starburst or irregular galaxies. This number confirms that the procedure works similarly in the BAT and in the control sample.

By applying also the completeness correction (equation 12), we derive that the total number of real mergers in the control sample is 8/0.8–10 which corresponds to a merger fraction of

$$f_{\mathrm{m,corr,control}} \simeq 10/247 \simeq 4.0^{+1.7}_{-1.2} \text{ per cent.} \qquad (15)$$

In addition, the large number of objects in this sample allows us to derive an estimate of P_{CASGM} which is more accurate than the one based on the BAT sample:

$$P_{CASGM} = (8^{+3.95}_{-2.77})/237 \simeq 0.034^{+0.017}_{-0.012}. \qquad (16)$$

Our results show that the average merger fraction of galaxies at redshift ~ 0 is very low, in accordance with the studies of Patton & Atfield (2008), Patton et al. (2002) and Koss et al. (2010), that claim a merger fraction of ~ 1–2 per cent. The higher value suggested by our work is probably related to a selection effect, because our control sample is not drawn as a random selection of galaxies in the pre-fixed redshift interval, but it is forced to follow the BAT sample's redshift distribution. This confirms the importance of building a control sample which reflects, as much as possible, all the key properties of the other sample. The merger fraction found in the control sample is significantly (3σ) lower than that found in the BAT sample.

4.2.2 The role of the galaxy mass distribution

The (stellar) mass distribution of galaxies hosting BAT AGN is very likely to be different from that of inactive galaxies or SDSS AGN (Koss et al. 2011), with BAT AGN typically residing in galaxies more massive than average.

The effects of galaxy mass upon the merger fraction measured through the *CASGM* method are uncertain.[14] However, if the mass dependence is relatively strong, we might obtain different merger fractions for the BAT and the control sample simply because of their different mass distributions. Therefore, it is necessary to check this hypothesis.

As a first step we evaluated galaxy stellar masses: this was done by converting the *ugriz* magnitudes from the SDSS into Johnson *BVRI* magnitudes (using the formulae in Blanton & Roweis 2007), calculating the distance modulus (DM) from the redshift of each galaxy (we assumed $H_0 = 71 \text{ km s}^{-1} \text{ Mpc}^{-1}$), and finally estimating the stellar masses as

$$\log(M_*/\mathrm{M_\odot}) = \log[\mathcal{M}_I(B-R)] + 0.4(I - \mathrm{DM} - \mathrm{I_\odot}), \qquad (17)$$

where $\mathcal{M}_I(B-R)$ is the mass to light ratio (in solar units) provided by Bell & de Jong (2001) for the I band, and as a function of

---

[14] Patton & Atfield (2008) find that the frequency of galaxy pairs is larger for low-luminosity (and, presumably, low-mass) than for high-luminosity galaxies; but this trend is reversed when they correct for perspective pairs. The *CASGM* method is somewhat in between the two cases: a galaxy pair which is well separated on the sky will be classified as a merger only if there are morphological anomalies (i.e. if the pair is physical); but the method cannot distinguish physical and perspective pairs if the sky separation is small. Then, the *CASGM*-measured merger fraction should have only a weak dependence on galaxy mass.
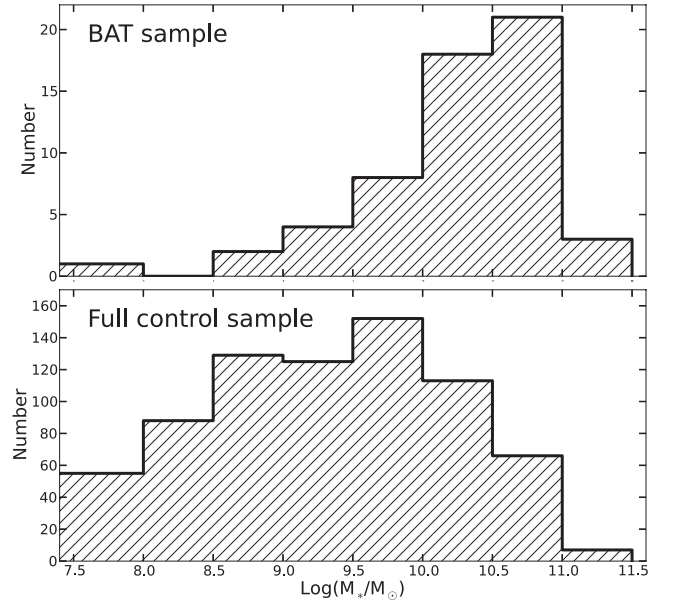

Figure 4. Comparison between the mass distributions of the BAT sample (top panel) and of the full control sample (bottom panel).

the $B - R$ colour of the galaxy; whereas I is the galaxy apparent I magnitude, and $\mathrm{I_\odot} = 4.52$ is the I absolute magnitude of the Sun.

Fig. 4 compares the distributions of stellar masses in the BAT and in the full control sample: the two distributions are quite different, as massive galaxies are much more frequent in the BAT sample. We note that this is partly caused by the contribution of the AGN within the galaxies of the BAT sample; however, the observed difference in luminosities is very large (the medians of the two samples differ by a factor of ~ 5), and cannot be explained in this way.

We checked whether this difference in the mass distributions could account for the difference in the merger fraction by building a mass-matched sample in the same way as we built a redshift-matched sample (see Section 2.2). In this case, we divided the galaxies in three mass bins ($M_*/\mathrm{M_\odot} < 10^{9.5}$; $10^{9.5} \leq M_*/\mathrm{M_\odot} < 10^{10.5}$; $M_*/\mathrm{M_\odot} \geq 10^{10.5}$), and extracted 173 systems from the full control sample.

Within the mass-matched control sample, 11 systems are classified as mergers by the *CASGM* combined criterion; this corresponds to an uncorrected merger fraction $f_{\mathrm{m,MMS}} = 11^{+4.4}_{-3.3}/173 \simeq 6.4^{+2.5}_{-1.9}$ per cent, and to a corrected merger fraction of $f_{\mathrm{m,corr,MMS}} = 3.9^{+3.3}_{-2.4}$ per cent, in very good agreement with the values for the redshift-matched control sample.

This result should be taken with caution, since the redshift distribution of the mass-matched control sample is different from that of the BAT sample. An ideal comparison should use a sample that *simultaneously* matches both the mass and redshift distributions of the BAT sample; unfortunately, our full control sample does not allow us to proceed in this way, as it includes only a small number (5) high-mass ($M_*/\mathrm{M_\odot} \geq 10^{10.5}$) systems at $z < 0.02$.

However, we can look at the simultaneous effect of both mass and redshift in two different ways. In the $0.02 \leq z < 0.03$ bin the full control sample includes a reasonable number (68) of high-mass systems; therefore, we extracted a mass-matched control sample within this redshift bin, where the combined criterion finds 15 mergers among the 162 systems. This corresponds to an uncorrected merger fraction $f_{\mathrm{m,MMS},z \geq 0.02} = 15^{+5.0}_{-3.8}/162 \simeq 9.3^{+3.1}_{-2.3}$ per cent, and to a

corrected merger fraction of $f_{\mathrm{m,corr,MMS},z\geq0.02} = 6.3^{+3.9}_{-3.0}$ per cent; both values are consistent with the results for the same redshift bin that we gave in Table 2 ($f_{\mathrm{m},z\geq0.02} = 7.7^{+3.3}_{-2.5}$ per cent, and $f_{\mathrm{m,corr},z\geq0.02} = 5.6^{+4.6}_{-3.2}$ per cent).

Instead, when looking at our full redshift range, we evaluate the uncorrected merger fraction in each bin of redshift and mass, and average them so as to reproduce the mass and redshift distribution of the BAT sample. In this way, we get an uncorrected merger fraction $f_{\mathrm{m,avg}} = 7.2^{+9.1}_{-2.7}$ per cent, and a corrected merger fraction[15] $f_{\mathrm{m,avg,corr}} = 3.7^{+11.5}_{-3.4}$ per cent. The large errors derive from the highly uncertain merger fractions of high-mass systems at $z <$ 0.02: if instead we make the very reasonable assumption that these are equal to what we find for high-mass systems at $0.02 \leq z < 0.03$ ($f_{\mathrm{m},z\geq0.02,\log(M)\geq10.5} = 8.8^{+5.3}_{-3.5}$ per cent, fully compatible both with the scarce high-mass data at $z < 0.02$, and with the redshift trend of the merger fractions in the other mass bins), we obtain an uncorrected merger fraction $f_{\mathrm{m,avg*}} = 5.9^{+3.5}_{-1.6}$ per cent, and a corrected one of $f_{\mathrm{m,avg*,corr}} = 2.5^{+4.6}_{-2.1}$ per cent.

We conclude that simultaneously controlling for the mass and redshift distributions cannot reconcile the merger fractions of the BAT and the control sample. This fact is proved (at the 1.8σ level) for the $0.02 \leq z < 0.03$ redshift bin. In the full sample it somewhat depends on the assumption that the merger fraction for galaxies with $M_* \geq 10^{10.5}$ M$_\odot$ does not change between $z = 0.003$ and 0.03: if such assumption is made, the (corrected) merger fractions of the two samples differ at the 2.6σ level.

5 SUMMARY AND CONCLUSIONS

In this work we focused on three main topics.

(i) *Software*. We have implemented the new software PYCASSO for the automated computation of the structural indexes of the *CASGM* system. Our procedures are entirely based on the definitions and relations presented in Conselice (2003) and Lotz et al. (2004), but we have implemented the possibility to use elliptical apertures, because they provide a better fit of the galaxy outline. Moreover, we carried on extensive tests on possible image degradations, so our software minimizes any data loss and smoothing effect and provides a stable and reliable analysis.

(ii) *Method*. We propose an improved technique for evaluating the merger fraction of a galaxy sample by means of the *CASGM* system. Indeed, we show that the original classification is biased towards irregular, edge-on and dusty galaxies, which tend to be misclassified as mergers. We propose a combined criterion between the A, S, G and M_{20} indexes, which leads to the complete blending of the *CAS* and *GM* methods and corrects nearly 80 per cent of the contamination. Then, we define the completeness and the reliability coefficients that allow a statistical correction of the merger fraction and further reduce possible residual errors in the automated classification.

(iii) *Application*. We have applied the *CASGM* analysis to a sample of local AGN host galaxies and a comparison sample, to extract their merger fractions and test whether there is an enhanced fraction of mergers among active galaxies. We found that in the BAT sample the merger fraction is 20^{+7}_{-5} per cent. In the redshift-matched control sample the merger fraction is $4.0^{+1.7}_{-1.2}$ per cent, and the difference is significant at the 3σ level. We obtain similar results for

a mass-matched control sample. Simultaneously matching redshift and mass leads to comparable but somewhat less significant results.

Our work is in agreement with other observational studies (Sanders et al. 1988; Koss et al. 2010) and numerical simulations (Barnes & Hernquist 1991; Di Matteo, Springel & Hernquist 2005; Hopkins et al. 2006) that suggest that galaxy interactions trigger the activity of the SMBH at their centre. The most likely scenario is that the strong gravitational perturbations drive large quantities of gas towards the centre of the remnant, originating both an intense starburst phase and an enhanced nuclear activity. Mergers may therefore be responsible not only for large-scale ($\sim10^3$ pc) distortions but also of the inflow of gas down to the typical scale of SMBH accretion ($\sim10^{-4}$ pc). Current numerical simulations cannot investigate entirely such a wide scale range, so observational studies have a key role for the comprehension of these phenomena. However, as we pointed out in Section 1, similar studies on higher redshift ($0.2 < z < 1.2$) galaxy samples (i.e. Pierce et al. 2007; Gabor et al. 2009; Cisternas et al. 2011) do not show any enhancements of the merger fraction of AGN host galaxies. Selection biases in the active galaxies sample and/or in the control sample could partially explain these contradicting results. For example, the aforementioned studies are based on other selection criteria (i.e. soft X-ray, 2–10 keV energy band), but, due to the significant fraction of obscured AGN (see Menci et al. 2008), they may detect a lower number of sources compared to our hard X-ray (15–195 keV) selection.

Therefore, while the results presented here and in previous observational studies (e.g. Koss et al. 2010) suggest that in the low-redshift ($z < 0.03$) Universe galaxy interactions trigger the activity of the SMBH at their centre, further researches that focus on an accurate and unbiased selection of galaxies – both at intermediate ($0.03 \leq z < 0.2$) and higher ($0.2 \leq z < 1.2$) redshifts – are mandatory to derive improved estimates on the occurrence and role of galaxy interactions on SMBH activity.

---

[15] If $f_{\mathrm{m}} \equiv N_{\mathrm{m}}/N_{\mathrm{sys}}$ is the uncorrected merger fraction, equation (14) implies that $f_{\mathrm{m,corr}} \equiv N_{\mathrm{m,real}}/N_{\mathrm{sys}} = (f_{\mathrm{m}} - P_{CASGM})/(C_{CASGM} - P_{CASGM})$.

REFERENCES

Abraham R. G., Tanvir N. R., Santiago B. X., Ellis R. S., Glazebrook K., van den Bergh S., 1996, MNRAS, 279, L47
Abraham R., van den Bergh S., Nair P., 2003, ApJ, 588, 218
Aihara H. et al., 2011, ApJS, 193, 29
Alonso M. S., Lambas D. G., Tissera P., Coldwell G., 2007, MNRAS, 375, 1017
Barnes J. E., 1988, ApJ, 331, 699
Barnes J. E., Hernquist L. E., 1991, ApJ, 370, 65
Barton E. J., Geller M. J., Kenyon S. J., 2000, ApJ, 530, 660
Bell E. F., de Jong R. S., 2001, ApJ, 550, 212
Bershady M., Jangren A., Conselice C., 2000, AJ, 119, 2645
Bertin E., Arnouts S., 1996, A&AS, 317, 393
Best P., Kaiser C. R., Heckman T. M., Kauffmann G., 2006, MNRAS, 368, L67
Blanton M. R., Roweis S., 2007, AJ, 133, 734
Blanton M. et al., 2003, ApJ, 594, 186
Bournaud F., Combes F., Jog C. J., Puerari I., 2005, A&A, 438, 570
Bournaud F. et al., 2011, ApJ, 730, 4
Byun Y. I., Freeman K. C., 1995, ApJ, 448, 563
Churazov E., Bruggen M., Kaiser C. R., Bohringer H., Forman W., 2001, ApJ, 554, 261
Cisternas M. et al., 2011, ApJ, 726, 57
Coldwell G. V., Lambas D. G., 2006, MNRAS, 371, 786
Conselice C., 2003, ApJS, 147, 1
Conselice C., 2006, ApJ, 638, 686
Conselice C. J., Bershady M. A., Gallagher J. S., 2000, A&A, 354, L21
Croton D. J. et al., 2006, MNRAS, 365, 11
Cusumano G. et al., 2010, A&A, 524, A64
Dahari O., 1984, AJ, 89, 966
Dahari O., 1985, ApJS, 57, 643
Darg D. W. et al., 2010, MNRAS, 401, 1552
Dekel A. et al., 2009, Nat, 457, 451
De Propris R., Conselice C. J., Liske J., Driver S. P., Patton D. R., Graham A. W., Allen P. D., 2007, ApJ, 666, 212
Di Matteo T., Springel V., Hernquist L., 2005, Nat, 433, 604
Dunlop J. S., McLure R. J., Kukula M. J., Baum S. A., O'Dea C. P., Hughes D. H., 2003, MNRAS, 340, 1095
Ellison S. L., Patton D. R., Simard L., McConnachie A. W., 2008, AJ, 135, 1877
Ellison S. L., Patton D. R., Mendel J. T., Scudder J. M., 2011, MNRAS, 418, 2043
Ferrarese L., Ford H., 2005, Space Sci. Rev., 116, 523
Ferrarese L., Merritt D., 2000, ApJ, 539, L9
Frei Z., Guhathakurta P., Gunn J., Tyson J. A., 1996, AJ, 111, 174
Fuentes-Williams Th., Stocke J. T., 1988, AJ, 96, 1235
Gabor J. M. et al., 2009, ApJ, 691, 705
Gebhardt K. et al., 2000, ApJ, 539, L13
Gehrels N., 1986, ApJ, 303, 336
Gehrels N. et al., 2004, ApJ, 611, 1005
Gerhard O. E., 1981, MNRAS, 197, 179
Gesch D. B., 2007, in Maune D., ed., Digital Elevation Model Technologies and Applications: The DEM Users Manual, 2nd edn. American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland, p. 99
Gesch D., Oimoen M., Greenlee S., Nelson C., Steuck M., Tyler D., 2002, J. Am. Soc. Photogramm. Remote Sens., 68, 5
Glasser G. J., 1962, J. Am. Stat. Assoc., 57, 648

Graham A. W., 2012a, ApJ, 746, 113
Graham A. W., 2012b, MNRAS, 422, 1586
Grogin N. A. et al., 2005, ApJ, 627, L97
Häring N., Rix H. W., 2004, ApJ, 604, L89
Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006, ApJS, 163, 1
Jogee S. et al., 2009, ApJ, 697, 1971
Kauffmann G., Haehnelt M., 2000, MNRAS, 311, 576
Keel W. C., Kennicutt R. C., Jr, Hummel E., van der Hulst J. M., 1985, AJ, 90, 708
Keres D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2
Kocevski D. D. et al., 2012, ApJ, 744, 148
Koss M., Mushotzky R., Veilleux S., Winter L., 2010, ApJ, 716, 125
Koss M. et al., 2011, ApJ, 739, 57
Koss M., Mushotzky R., Treister E., Veilleux S., Vasudevan R., Trippe M., 2012, ApJ, 746, 22
Koulouridis E., Plionis M., Chavushyan V., Dultzin-Hacyan D., Krongold Y., Goudis C., 2006, ApJ, 639, 37
Lake G., Dressler A., 1986, ApJ, 310, 605
Le Févre et al., 2000, MNRAS, 311, 565
Li C., Kauffmann G., Heckman T. M., White S. D. M., Jing Y. P., 2008, MNRAS, 385, 1915
Liu X., Shen Y., Strauss M. A., 2012, ApJ, 745, 94
Lotz J. M., Primack J., Madau P., 2004, AJ, 128, 163
Lotz J. M., Jonsson P., Cox T. J., Primack I. R., 2008, MNRAS, 391, 1137
McNamara B. R., Nulsen P. E. J., 2007, ARA&A, 45, 117
Magorrian J. et al., 1998, AJ, 115, 2285
Mapelli M., Moore B., Bland-Hawthorn J., 2008, MNRAS, 388, 697
Marconi A., Hunt L. K., 2003, ApJ, 589, L21
Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, MNRAS, 351, 169
Menci N., Fiore F., Puccetti S., Cavaliere A., 2008, ApJ, 686, 219
Merloni A., Heinz S., 2012, preprint (arXiv:1204.4265)
Miller R. H., Smith B. F., 1980, ApJ, 235, 421
Miller Ch. J., Nichol R. C., Gómez P. L., Hopkins A. M., Bernardi M., 2003, ApJ, 597, 142
Mirabel I. F., 2001, Astrophys. Space Sci. Suppl., 277, 371
Negroponte J., White S. D. M., 1983, MNRAS, 205, 1009
Patton D. R., Atfield J. E., 2008, ApJ, 685, 235
Patton D. R., Carlberg R. G., Marzke R. O., Pritchet C. J., da Costa L. N., Pellegrini P. S., 2000, ApJ, 536, 153
Patton D. R. et al., 2002, ApJ, 565, 208
Paturel G., Petit C., Prugniel Ph., Theureau G., Rousseau J., Brouty M., Dubois P., Cambrésy L., 2003, A&A, 412, 45
Peng C. Y., Ho L. C., Impey C. D., Rix H. R., 2002, AJ, 124, 266
Petrosian V., 1976, ApJ, 209, L1
Petrosian A. R., 1982, Afz, 18, 548
Pierce C. M. et al., 2007, ApJ, 660, 19
Rafanelli P., Violato M., Baruffolo A., 1995, AJ, 109, 1546
Ramos Almeida C., Tadhunter C. N., Inskip K. J., Morganti R., Holt J., Dicken D., 2011, MNRAS, 410, 1550
Sales L. V., Navarro J. F., Theuns T., Schaye J., White S. D. M., Frenk C. S., Crain R. A., Dalla Vecchia C., 2012, MNRAS, 423, 1544
Sancisi R., Fraternali F., Oosterloo T., van der Hulst Th., 2008, A&AR, 15, 189
Sanders D. B., Soifer B. T., Elias J. H., Madore B. F., Matthews K., Neugebauer G., Scoville N. Z., 1988, ApJ, 325, 74
Scarlata C. et al., 2007, ApJS, 172, 494
Schawinski K. et al., 2006, Nat, 442, 888
Schawinski K., Thomas D., Sarzi M., Maraston C., Kaviraj S., Joo S. J., Yi S. K., Silk J., 2007, MNRAS, 382, 1415
Schawinski K. et al., 2009, ApJ, 690, 1672
Schawinski K., Dowlin N., Thomas D., Urry C. M., Edmondson E., 2010, ApJ, 714, L108
Schmitt H. R., 2001, AJ, 122, 2243
Serber W., Bahcall N., Ménard B., Richards G., 2006, ApJ, 643, 68
Silverman J. D. et al., 2011, ApJ, 743, 2
Smirnova A. A., Moiseev A. V., Afanasiev V. L., 2010, MNRAS, 408, 400

Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776

Vikram V., Wadadekar Y., Kembhavi A. K., Vijayagovindan G. V., 2010, MNRAS, 409, 1379

Virani S. N., De Robertis M. M., VanDalfsen M. L., 2000, AJ, 120, 1739

Volonteri M., Bellovary J., 2012, Rep. Progress Phys., 75, 124901

Waskett T. J., Eales S. A., Gear W. K., McCracken H. J., Lilly S., Brodwin M., 2005, MNRAS, 363, 801

White S. D. M., 1978, MNRAS, 184, 185

APPENDIX A: DATA PROCESSING ALGORITHMS

The image analysis process is split into three main phases: data acquisition, pre-processing and processing. In the first two phases we essentially use publically available codes (MONTAGE and SExtractor), whereas for the processing phase we developed the software PYCASSO... (*Appendix is presented in its entirety in the electronic edition.*)

APPENDIX B: IMAGE DEGRADATION

Quantitative analysis can be distorted even by small image degradations. Therefore, it is important to use procedures that minimize image alterations. The most common image-altering operations performed within the *CASGM* analysis are translations and rotations... (*Appendix is presented in its entirety in the electronic edition.*)

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix A. Data processing algorithms.
Appendix B. Image degradation (http://mnras.oxfordjournals.org/lookup/suppl/doi:10.1093/mnras/stt358/-/DC1).

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a TEX/LATEX file prepared by the author.