Compact groups from the Millennium Simulations – I. Their nature and the completeness of the Hickson sample

Eugenia Díaz-Giménez$^{1,2}$ and Gary A. Mamon$^{3,4}$

$^1$Instituto de Astronomía Teórica y Experimental, IATE, Observatorio Astronómico, Laprida 854, Córdoba, Argentina
$^2$Consejo de Investigaciones Científicas y Técnicas de la República Argentina (CONICET)
$^3$Institut d’Astrophysique de Paris (UMR 7095: CNRS & Univ. Pierre & Marie Curie) 98 bis Bd. Arago, F–75014 Paris, France
$^4$Astrophysics & BIPAC, Department of Physics, University of Oxford, Oxford OX1 3RH

Accepted 2010 July 19. Received 2010 July 18; in original form 2010 May 20

ABSTRACT
We identify compact groups of galaxies (CGs) within mock galaxy catalogues from the Millennium Simulation at $z = 0$ with the semi-analytic models (SAMs) of galaxy formation of Bower et al., Croton et al. and De Lucia & Blaizot. CGs are identified using the same 2D criteria as those visually applied by Hickson (1982) to his CGs (HCGs), but with a brightest galaxy magnitude limit, and we also add the important effect of observers blending close projected pairs. Half of the mock CGs identified in projection contain at least four accordant velocities ($mvCGs$), versus 70 per cent for HCGs. In comparison to $mvCGs$, the HCGs are only 8 per cent complete at distances $<9000 \text{ km s}^{-1}$, missing the CGs with small angular sizes, a strongly dominant galaxy, and (for the second SAM) the $mvCGs$ that are fainter and those with lower surface brightness. 10 per cent of the mock $mvCGs$ are identical to the parent virialized group, meaning that they are isolated, while the remainder are embedded in their parent virialized groups. We explore different ways to determine the fraction of physically dense groups given the data from the simulations. Binding energy criteria turn out to be inapplicable given the segregation between galaxies and dark matter particles. We rely instead on the combination of the three-dimensional length of the CGs (maximum real space galaxy separation) and their elongation along the line of sight (ratio of maximum line-of-sight to maximum projected separations), restricting ourselves in both cases to the smallest quartets within the CGs. We find that between 64 and 80 per cent (depending on the SAM) of the $mvCGs$ have 3D lengths shorter than $200 h^{-1} \text{ kpc}$, between 71 and 80 per cent have line-of-sight elongations less than 2, while between 59 and 76 per cent have either 3D lengths shorter than $100 h^{-1} \text{ kpc}$ or both lengths shorter than $200 h^{-1} \text{ kpc}$ and elongations smaller than 2. Therefore, chance alignments (CAs) of galaxies concern at most 40 per cent of the $mvCGs$. These CAs are mostly produced from larger host groups, but a few have galaxies extending a few Mpc beyond the host group. The $mvCGs$ built with the Hickson selection (respectively without the close projected pair blending criterion) have 10 per cent higher (lower) fractions of physically dense systems.

Key words: methods: data analysis – galaxies: clusters: general – galaxies: distances and redshifts.

1 INTRODUCTION

Compact groups (CGs) are small, relatively isolated systems of typically four or five luminous galaxies in close proximity to one another. The first example of a CG was found by Stephan (1877). Several catalogues of CGs are now available: Rose (1977) and Hickson (1982) visually identified CGs on POSS I photographic plates. After the Hickson compact group (HCG) catalogue, several CG catalogues have been automatically extracted from galaxy catalogues, themselves automatically extracted from photographic plates: from the COSMOS/UKST Southern Galaxy Catalog (Pandorini, Iovino & MacGillivray 1994; Iovino 2002), from the DPOSS catalogue (Iovino et al. 2003; de Carvalho et al. 2005) or CCD frames from the Sloan Digital Sky Survey (SDSS) photometric catalogue (Lee et al. 2004). CG catalogues have also been extracted from galaxy catalogues in redshift space: from the CfA2 (Barton et al. 1996), Las Campanas (Allam & Tucker 2000), and SDSS...
(Deng et al. 2007; McConnachie et al. 2009) surveys, as well as from the 3D UZC Galaxy Catalog (Focardi & Kelm 2002). CGs are so compact that the median projected galaxy separation in HCGs is only 39 h⁻¹ kpc (Hickson et al. 1992).

HCGs have been studied in detail, in particular their internal structures, shapes, morphologies, luminosities and environments (Hickson 1982; Hickson et al. 1984; Mamon 1986; Hickson, Kindl & Huchra 1988; Hickson & Rood 1988; Mendes de Oliveira & Hickson 1991; Zepf 1993; Moles et al. 1994; Prandoni et al. 1994; Kelm & Focardi 2004; Tovmassian, Plionis & Torres-Papaqui 2006). To summarize, these studies indicate that CG galaxies have star formation properties, colours and morphological mixes that lie in between binary galaxies and isolated ones.

The nature of the CGs has been a puzzling matter for quite some time. How can a few bright galaxies coexist within less than 100 kpc, given that galaxies are expected to merge fast in such systems (Carnevali, Cavaliere & Santangelo 1981; Barnes 1985; Mamon 1987; Bode, Cohn & Lugger 1993)? There are three schools of thought on this matter. One view is that compact groups are recently formed dense systems that are about to coalesce into a single galaxy (Hickson & Rood 1988). The galaxies lost in the merger may be replenished by galaxies in the loose group environment (Diaz, Geller & Ramella 1994), and the predicted rate of formation of CGs appears to be sufficient to explain the observed frequency of HCGs (Mamon 2000). The second view states that CGs may be transient unbound cores of looser groups (Rose 1977; Ramella et al. 1994; Tovmassian, Yam & Tiersch 2001). And the third scenario places CGs as chance alignments of galaxies along the line of sight within larger loose groups (Rose 1977 for CGs elongated in projection; Mamon 1986 and Walke & Mamon 1989 in general), clusters (Walke & Mamon 1989) and cosmological filaments (Hernquist, Katz & Weinberg 1995). In this scenario, the numerous signs of interaction and star formation is explained by the frequent occurrence of binaries and triplets in the chance alignments (Mamon 1992).

If CGs are physically dense, their dynamical times should be short (1 per cent of the age of the Universe), and the hot intragroup gas should trace a smooth gravitational potential. The launch of X-ray observatories with good sensitivity in the soft X-ray band (ROSAT, ASCA, Chandra and XM–Newton) has led to the detection of hot X-ray-emitting gas from many CGs. Since the X-ray emissivity scales as the square of the gas density, X-ray emission is less prone to projection effects than optical surveys (but see Ostriker, Lubin & Hernquist 1995). However, although 22 HCGs were detected out of 32 pointed observations (Ponman et al. 1996), it is not clear what the global fraction of detections would be on the full sample of HCGs (69 [92] groups with at least four [3] accordant velocities, according to Hickson et al. 1992). Moreover, some of the detected groups appear clumpy (e.g. HCG 16 according to Santos & Mamon 1999), which strongly suggests their unvirialized state.

The distinction between compact groups that are dense in 3D, and chance alignments within loose groups or larger filaments is difficult, because redshift space distortion introduces uncertainties in the computation of the line-of-sight coordinate which might result in misidentified compact configurations. For a group with line-of-sight velocity dispersion $\sigma_v$, the redshift distortion will amount to a spread of $\delta r = \sigma_v / c_H$ in the line-of-sight coordinate. Assuming that the squared velocity dispersion is half the squared circular velocity at the virial radius, $\sigma_v^2 = (1/2) GM(r_v)/r_v$ (appropriate for an $\rho \propto 1/r^2$ density profile), one finds $\delta r / r_v = (1/2) \sqrt{100} = 5$ if the virial radius is defined where the mean density at that radius is 100 times the critical density of the Universe. So redshift space distortions prevent measuring distances within virialized systems (see also the Introduction of Walke & Mamon 1989).

Nevertheless, there is one CG meeting the HCG criteria discovered by one of us (Mamon 1989) that is so close (within the Virgo cluster) that surface brightness fluctuation distance measurements by Mei et al. (2007) are able to settle the issue of its nature: Mamon (2008) concludes that this CG is a chance alignment of galaxies along the line of sight, at least 440 kpc and most probably 2 Mpc long.

In summary, even though many efforts have been devoted to look for an explanation about the nature of CGs, the debate is still wide open.

The advent of increasingly realistic cosmological simulations now allow one to distinguish whether CGs are truly dense in 3D, or caused by chance alignments within looser groups, filamentary structures or the general field. In an early pioneering attempt, Hernquist et al. 1995 identified galaxies as dense knots of cold gas in their N-body + hydrodynamical simulation, and searched for CGs in redshift space in many viewing directions. They found four CGs with at least four accordant velocities, all of which were longer than 2 $h^{-1}$ Mpc along the line of sight (one was as long as 4 $h^{-1}$ Mpc), and yet presented accordant velocities, despite the (Hubble law) stretching of velocities caused by their elongation along the line of sight. The analysis of Hernquist et al. suffers from several drawbacks (according to present-day standards for cosmological simulations): the simulation box was small (44 $h^{-1}$ Mpc wide); the mass resolution was poor (their simulation had 32 dark matter particles and 32 gas particles, and their galaxies were identified with as few as eight gas particles); and the spatial resolution was poor (the dark matter particles had a softening length of 10 $h^{-1}$ kpc). Furthermore, the identification of galaxies with knots of dense gas was not optimal, especially that feedback from supernovae and active galactic nuclei were not incorporated.

In this work, we quantify the fraction of CGs that can be considered as physically dense entities in samples of automatically identified CGs, based upon more realistic cosmological N body simulations. At present, one can build realistic CGs in two ways: (1) from dissipationless cosmological simulations, on top of which galaxies are painted using fairly complex semi-analytical galaxy formation/evolution models; (2) from hydrodynamical codes that resolve galaxies. We have chosen the first approach and use for this purpose the largest cosmological N body simulation ever performed (in 2006, when the present study began), the Millennium Run (Springel et al. 2005), on which galaxies were identified in three ways, using three different state-of-the-art semi-analytic models (SAMs) of galaxy formation (Bower et al. 2006; Croton et al. 2006; De Lucia & Blaizot 2007).

These three galaxy samples provide an opportunity to test both, projection effects and the real nature of systems identified using standard algorithms like that proposed by Hickson (1982). The CGs are identified in mock redshift-space catalogues constructed from the real-space galaxy sample derived with the semi-analytical model, from the Millennium Run.

In comparison with the analysis of Hernquist et al., our study is based upon a simulation in a box whose volume is over 1 million times greater, with 30 thousand times as many particles, 25 times finer mass resolution and a softening scale four times smaller. However, the simulation we use does not contain gas particles, so the galaxy parameters are highly dependent on the physics of galaxy formation and evolution of the three SAMs that we analyse.

We focus here on the HCG catalogue, which is by far the best-studied sample of compact groups.
2 CONSTRUCTION AND CLASSIFICATION OF THE COMPACT GROUP SAMPLE

2.1 Observed compact group sample

We use the HCG catalogue of compact groups, with photometry measured by Hickson, Kindl & Auman (1989) in the B and (presumably Johnson) R bands. Hickson (1982) found 100 HCGs, and Hickson et al. (1992), who measured the redshifts for virtually all galaxies, built a velocity sample of vHCGs by eliminating galaxies lying at more than 1000 km s\(^{-1}\) from the group’s median velocity. In this manner, they obtained 92 HCGs with at least three accordant velocities and 69 HCGs with at least four accordant velocities. We extracted the photometry and velocities using the table VII/213/galaxies in VizieR\(^1\) (Ochsenbein, Bauer & Marcout 2000). This data base contains the velocities for all galaxies except six. We found the redshifts for these six galaxies (Table 1) in the NASA/IPAC Extragalactic Data base (NED).\(^2\)

| Galaxy | \(v\) (km s\(^{-1}\)) | \(\epsilon(v)\) (km s\(^{-1}\)) | Reference |
|--------|----------------------|----------------------|-----------|
| 18b    | 4105                 | 25                   | Falco et al. (1999) |
| 19c    | 4253                 | 23                   | de Carvalho et al. (1997) |
| 19d    | 20443                | 26                   | de Carvalho et al. (1997) |
| 51g    | 7532                 | 41                   | de Vaucouleurs et al. (1991) |
| 57h    | 9240                 | 105                  | Hickson (1993), Barton, de Carvalho & Geller (1998) |
| 100d   | 5590                 | 32                   | Hickson (1993) |

Notes. The radial heliocentric velocities and their errors are given in columns 2 and 3, respectively. For HCG57h, the velocity is from the first reference, while the error is from the second.

The layout of this paper is as follows. In Section 2, we present the different steps for the construction of the mock CG catalogue and Section 3 describes how the resulting CGs are classified. The conclusions are summarized and discussed in Section 4. Once our analysis was well advanced, we learnt about the work of McConnachie, Ellison & Patton (2008), who performed a similar analysis of the properties of Hickson-like CGs from the De Lucia & Blaizot (2007) galaxy catalogue, and found that 70 per cent of the mock CGs selected in projection were caused by chance alignments of galaxies. We highlight in Section 4.2 the similarities and several important differences between our two studies.

2.2 Basic scheme for mock compact group samples

Our mock catalogues of CGs are built in several steps, in which we

(i) simulate the gravitational evolution of a large piece of the Universe, represented by collisionless (dark matter) particles;

(ii) attach galaxies to the simulation with a semi-analytical galaxy formation model

(iii) convert to a mock galaxy catalogue in redshift space;

(iv) convert to a mock 2D CG catalogue (hereafter mpCG for mock CG in projection), by applying the HCG selection criteria;

(v) convert the mpCG catalogue to a velocity-filtered mock CG catalogue (hereafter, mvCG for mock velocity-filtered CG), by removing galaxies with discordant redshifts;

(vi) convert the mvCG catalogue to a mock velocity-filtered HCG catalogue (hereafter, mvHCG for mock velocity-accordant Hickson compact group), by randomly selecting groups according to the completeness of the HCG as a function of group surface brightness, brightest galaxy magnitude and its contribution to the total group luminosity.

The last step is motivated by the strong incompleteness of the HCG catalogue in surface magnitude and galaxy magnitude (see Section 2.7, below).

A list of the different acronyms used to refer to the different samples is provided in Table 2.

2.3 Dark matter particle simulation

We use the Millennium Simulation, which is a cosmological Tree-Particle-Mesh (TPM, Xu 1995) N-body simulation (Springel et al. 2005), which evolves 10 billion (2\(^{160}\)) dark matter particles in a 500 h\(^{-1}\) Mpc periodic box, using a comoving softening length of 5 h\(^{-1}\) kpc.\(^3\) The cosmological parameters of this simulation

\[^{3}\] We have omitted group HCG 54, which is the HCG with the smallest projected radius, as it appears to be either a group of H\(\alpha\) regions in a single galaxy (Arkhipova et al. 1981) or the end result of the merger of two disc galaxies (Verdes-Montenegro et al. 2002).

\[^{4}\] This is also clear in the \(B_T\) and \(B_T^3\) magnitudes given in Hickson et al. 1989, although this was not discussed by these authors, but was also noted by Sulentic (1997).

\[^{5}\] The Millennium Simulation, run by the Virgo Consortium, is publicly available at http://www.mpa-garching.mpg.de/millennium

1 http://webviz.u-strasbg.fr/viz-bin/VizieR
2 http://nedwww.ipac.caltech.edu

Table 1. Additional HCG galaxy redshifts found in NED.

| Galaxy | \(v\) (km s\(^{-1}\)) | \(\epsilon(v)\) (km s\(^{-1}\)) | Reference                           |
|--------|----------------------|----------------------|-----------------------------------|
| 18b    | 4105                 | 25                   | Falco et al. (1999)              |
| 19c    | 4253                 | 23                   | de Carvalho et al. (1997)        |
| 19d    | 20443                | 26                   | de Carvalho et al. (1997)        |
| 51g    | 7532                 | 41                   | de Vaucouleurs et al. (1991)     |
| 57h    | 9240                 | 105                  | Hickson (1993), Barton, de Carvalho & Geller (1998) |
| 100d   | 5590                 | 32                   | Hickson (1993)                  |
correspond to a flat cosmological model with a non-vanishing cosmological constant ($\Lambda$CDM): $\Omega_m = 0.25$, $\Omega_b = 0.75$, $\sigma_8 = 0.9$ and $\Omega = 0.73$. The simulation was started at $z = 127$, with the particles initially positioned by displacing particles initially in a glass-like distribution according to the $\Lambda$CDM primordial density fluctuation power spectrum. The $10^9$ particles of mass $8.6 \times 10^5 h^{-1} M_{\odot}$ are then advanced with the TPM code, using 11,000 internal time-steps, on a 512-processor supercomputer. The positions and velocities of the 10 billion particles were saved at 64 epochs (leading to nearly 20 TB of data).

### 2.4 Modelling galaxies

We consider the $z = 0$ outputs from three different SAMs of galaxy formation by Bower et al. (2006), Croton et al. (2006) and De Lucia & Blaizot (2007), (B06, C06 and DLB, respectively), where each model was applied in turn to the outputs of the Millennium Simulation described above. Note that, while the B06 and C06 models were developed independently, the DLB model is essentially the same as the C06 model, except that the merger rate is reduced by a factor of 2, the magnitudes are derived using spectral synthesis models based upon a different initial mass function (Chabrier 2003 instead of Salpeter 1955) with fewer low mass stars and the treatment of radiative transfer to dust is much more refined.

The three SAMs produce galaxy positions, velocities, as well as absolute magnitudes (in five or more optical and near-infrared wavebands, all including Johnson R), as well as other quantities. To summarize, the branches of the halo merger tree (produced by the Millennium Simulation) are followed forward in time, and the following astrophysical processes are applied: gas infall and cooling, early reheating of the intergalactic medium by photoionization, star formation, black hole growth, AGN and supernova feedback, galaxy mergers, spectrophotometric evolution, etc. The model parameters have been adjusted to produce a good match to the observed properties of local galaxies. In these SAMs, AGN feedback is responsible for the absence of cooling flows in rich clusters, for the cut-off at the bright end of the galaxy luminosity function and for the number density properties of the most massive galaxies at all redshifts. Also, the early reheating of the IGM by photoionization is responsible for suppressing gas cooling in haloes below a circular velocity that is independent of redshift (or nearly so). The three SAMs produce $z = 0$ galaxy luminosity functions that are in good agreement with observations in both the $b_I$ and $K$ wavebands, with an excess of galaxies at very bright luminosities for all three models and a slight excess at faint luminosities for the C06 and DLB models. Moreover, the B06 SAM provides several other observational predictions: the $b_I$ and $K$ galaxy luminosity functions at higher redshifts, the global history of star formation and the local black hole mass versus bulge mass relation.

All three SAMs produce around 10 million galaxies at $z = 0$. The galaxy samples appear to be complete at least to $M_R = 5 \log h < -17.4$ with stellar masses $M_* > 10^9 h^{-1} M_{\odot}$ (C06) or $M_* > 3 \times 10^9 h^{-1} M_{\odot}$ (B06, DLB).

Each of the three SAMs has its strengths and weaknesses. The B06 model computes galaxy mergers by inferring the positions of galaxies in their halo through typical values of their energies and angular momentum in units dimensionless to the virial scales of the haloes. In contrast, the C06 and DLB models have the advantage of estimating the merger rates directly from the positions of subhaloes in the dark matter simulation. They both use the same analytical formula for the orbital decay time by dynamical friction once the subhalo masses fall below their resolution limit, where the DLB time is twice the C06 time, which itself matches almost perfectly the decay time that Jiang et al. (2008) calibrated on high-resolution cosmological hydrodynamical simulations. Unfortunately, C06 do not provide the galaxy merger trees, so it is difficult to derive the history of star formation of a given galaxy. Also, while B06 find a Red Sequence with increasing red colours for increasingly higher stellar masses, the C06 catalogue shows a colour–luminosity relation for the Red Sequence galaxies that flattens at high luminosity, contrary to observations, and a similar effect is seen in the DLB galaxy output (as shown by Bertone et al. 2007). Still, the B06 colours are too blue and fit somewhat less well the SDSS colour distribution than do the DLB colours (Mateus, Jimenez & Gaztañaga 2008). The SAM of Cattaneo et al. (2006) reproduces better the colours of galaxies, but its output is not public and the galaxy positions are determined stochastically (like B06) rather than by following the dark matter subhaloes (like C06 and DLB). The DLB catalogue produces galaxies whose present-day small-scale segregation of recently formed stellar mass is too large, while that of B06 matches well that observed with the SDSS (Mateus et al.). This is surprising given that the B06 model treats galaxy mergers using stochastic

---

### Table 2. List of acronyms used throughout this work.

| Acronym | Description |
|---------|-------------|
| CG | General compact groups |
| HCG | Hickson compact groups |
| pHCG | HCGs that strictly meet the Hickson (1982) criteria and $R_b \leq 14.44$ |
| vHCG | Velocity-accordant pHCGs |
| pmpCG | Particle mock projected compact groups, which strictly meet the Hickson (1982) criteria and $R_b \leq 14.44$ |
| mpCG | Observable mock projected compact groups (same as pmpCG, but accounting for galaxy confusion) |
| pmvCG | Particle mock velocity-accordant compact groups |
| mvCG | Observable mock velocity-accordant compact groups (same as pmvCG, but accounting for galaxy confusion) |
| mvHCG | Observable mock velocity-accordant compact groups with Hickson’s biases |
| CA | Chance alignment of galaxies |
| CALG | Chance alignment of galaxies within looser groups |
| CAF | Chance alignment of galaxies within filaments |
| PG3D | Parent groups identified in real space |

---

6 The $z = 0$ luminosity function of the De Lucia & Blaizot (2007) model is given by Bertone, De Lucia & Thomas (2007).
positions rather than the positions of the subhaloes with which the galaxies are associated (see above). However, the DLB model predicts a little better than B06 the analogous segregation for intermediate age (0.2–0.5 Gyr) stellar mass (Mateus et al.). But the present-day galaxy-merger rate of DLB appears too low, while that of B06 matches well the observations of the frequency of galaxy pairs (Mateus 2008).

In summary, it is very difficult to decide which of the three SAMs is most appropriate for our study of CGs, and we therefore decided to analyse the outputs of all three of them. We will find and illustrate several important differences in the properties of mock CGs predicted from these three models.

2.5 Mock galaxy catalogues

Using the snapshots at \( z = 0 \), we construct mock catalogues in redshift space. For the three SAMs, we obtain redshifts by adding the Hubble flow to the peculiar velocities projected in the line-of-sight direction. We compute the observer-frame galaxy apparent magnitudes from the rest-frame absolute magnitudes provided by the semi-analytical model. These apparent magnitudes are converted to the observer frame using tabulated \( k + e \) corrections (Poggianti 1997).

Our mock catalogue is constructed by viewing the full volume of the simulation box from one of its eight vertexes (\( \xi_{\text{max}} \sim 0.17, \pi/2 \text{sr} = 5156 \text{deg}^2 \) ). We set an apparent magnitude limit \( R = 17.44 \), equal to the limit we set on the HCG groups to match the SDSS spectroscopic catalogue for later comparisons (see Section 2.1).

In order to increase the statistical significance of our results, we considered eight observers situated at the eight vertexes of the simulation box, and all the identification procedures were performed on these eight samples to finally combine the resulting CGs into one larger sample. These eight samples are almost fully statistically independent, since 3/4 of the mock CGs selected in the cone (see Fig. 3) lie within half the box size (256 \( h^{-1} \text{Mpc} \)) (see upper right-hand panel of Fig. 2). Table 3 summarizes the main properties of the three mock galaxy catalogues seen from one of its vertexes.

The completeness of our magnitude+volume limited mock catalogues might be an important issue that could bias the results. The implications on the results of using magnitude+volume limited samples will be carefully tested in Section 2.8.

We identify regular groups of galaxies in the simulation box by applying a Friends-of-Friends (FoF) algorithm in real space (Davis et al. 1985) to the galaxies. We adopt a linking length of \( l = 0.17 n^{-1/3} \), where \( n \) is the mean space density of galaxies. The factor 0.17 roughly corresponds to an overdensity of 100 relative to the critical density of the Universe, roughly the minimum overdensity (hence maximum radius) where cosmological structures are in dynamical equilibrium (Bryan & Norman 1998, but recent work by Cuesta et al. 2008 shows that on the mass scales of groups, the radius of equilibrium is roughly 30 per cent greater). We denote these groups the PG3Ds for Parent groups selected in real space, and will later check if the mock CGs extend beyond these PG3Ds.

### Table 3. Mock galaxy catalogues (\( R < 17.44 \)).

| Mock          | #    | \( z_{\text{med}} \) | \( \pi_{\text{red}} \) |
|--------------|------|----------------------|----------------------|
| Bower et al. (2006) | 556224 | 0.0998 | 0.062 |
| Croton et al. (2006) | 446153 | 0.0969 | 0.059 |
| De Lucia & Blaizot (2007) | 1034619 | 0.1114 | 0.089 |

Notes: #: number of galaxies seen from a single vertex of the simulation box, \( z_{\text{med}} \): median redshift, \( \pi_{\text{red}} \): space number density within 90 \( h^{-1} \text{Mpc} \), defined in equation (1).

2.6 Mock compact groups selected in projected space

In this work, we use an automated \textit{mpCG} search algorithm very similar to that described by Hickson (1982), applied to the three mock galaxy catalogues. The algorithm defines as \textit{mpCGs} those systems that satisfy the following conditions:

1. \( 4 \leq N \leq 10 \) (population)
2. \( \mu_R < 26 \text{ mag arcsec}^{-2} \) (compactness)
3. \( \theta_N > 3 \theta_G \) (isolation)
4. \( R_{\text{brightest}} \leq 14.44 \) (flux limit),

where \( N \) is the total number of galaxies whose \( R \)-band magnitude satisfies \( R < R_{\text{brightest}} + 3 \), where \( R_{\text{brightest}} \) is the magnitude of the brightest galaxy; \( \mu_R \) is the mean \( R \)-band surface magnitude, averaged over the smallest circle circumscribing the galaxy centres, \( \theta_G \) is the angular diameter of this smallest circumscribed circle; \( \theta_N \) is the angular diameter of the largest concentric circle that contains no other galaxies within this magnitude range or brighter.

Note that the fourth criterion (which implies \( R_{\text{lim}} \leq 17.44 \)) was not considered by Hickson (1982). This restriction is fundamental for avoiding selection biases, as will be demonstrated in Section 2.8.

The main steps of this algorithm are summarized in the flowchart of Fig. 1.

Now, some CGs meeting Hickson’s criteria might be embedded within larger CGs that also meet Hickson’s criteria (with larger isolation annuli). For such groups, we thus have two choices for our CG selection algorithm: select the smaller (sub-)group (solid portion only in the flowchart of Fig. 1) or the larger group (with the dashed portion of the flowchart of Fig. 1). The percentages of CGs containing smaller CGs are 13, 10 and 6 per cent for B06, C06 and DLB models, respectively. The HCG sample was selected according to the larger group (P. Hickson, private communication).

However, of the 100 groups in the original HCG sample, only one has a definite subgroup (HCG 17). Therefore, it is not clear that P. Hickson always followed the larger group algorithm. In what follows, we adopt the larger group algorithm (i.e. including dashed portion only in the flowchart of Fig. 1). However, our results turn out to depend little on the choice among these two algorithms.

To accelerate this algorithm, we have used the subroutines of the \texttt{healpix}\(^7\) package to find neighbours, and the \texttt{stripack}\(^8\) subroutines to compute the centres and radii of the minimum circles. Given that our mock catalogues have edges (the limits of the cone), we discarded CGs lying near the edges since those groups will be fictitiously isolated. Then, we kept with a safe sample of CGs that lies in the range \( \alpha > 5^\circ \) & \( \alpha < 85^\circ \), and \( \delta > 5^\circ \) & \( \delta < 85^\circ \) (solid angle \( \Delta \Omega = 1.2693 \text{ sr} \)).

Using this algorithm, we find 7580, 4756 and 15383 mock CGs in the B06, C06 and DLB samples, respectively.

Now, the galaxies in the mock galaxy catalogues are simply point particles. However, when one observes two galaxies that lie so close in projection on the plane of the sky that their isophotes overlap, they

\(^7\) http://healpix.jpl.nasa.gov/index.shtml
\(^8\) http://people.sc.fsu.edu/~burkardt/f_\texttt{arc}/stripack/stripack.html
risk being blended into a single object. This galaxy confusion can be important for CGs, which by definition often have overlapping isophotes. For example, observed CG catalogues should have fewer very dense groups than mock CG catalogues. We therefore included one extra observability criterion: two galaxies are confused and blended if their projected separation is smaller than the sum of their half light radii, in which case we sum their luminosities and adopt their absolute magnitude in the $R$-band, according to Shen et al. (2003, equations 14 and 15 therein).

Hereafter (see Table 2), we refer to these observable mock projected compact groups as $mpCG$s and denote the original ones by $pmpCG$s (for particle-mock-projected compact groups). The $mpCG$s are built with the Hickson criteria given at the beginning of Section 2.6 and thus contain at least four galaxies (after the pair-blending procedure). We understand that our pair-blending criterion is simplistic and may be somewhat liberal in defining confused galaxy pairs. In reality, the observed CGs should lie in between our $pmpCG$s and our $mpCG$s, but probably much closer to the $mpCG$s. We therefore adopt the observable criterion (hence, the $mpCG$s) in what follows, unless explicitly stated otherwise.

### 2.7 Mock compact groups after velocity filtering

We then built a sample of velocity-filtered mock compact groups on top of the respective $pmpCG$ and $mpCG$ samples, which we call the $pmvCG$ (particle-mock-velocity-filtered compact group) and $mvCG$ (mock velocity-filtered compact group) samples (see Table 2) with the following iterative procedure (see Hickson et al. 1992):

1. compute the median velocity of the group, $v_{\text{median}}$;
2. discard those galaxies with $|v - v_{\text{median}}| > 1000\ k\mathrm{km\ s}^{-1}$;
3. if at least $n_{\text{min}}$ galaxies remain, iterate until no galaxies are dropped or the group disappears ($n < n_{\text{min}}$);
4. save those CGs that have at least $n_{\text{min}}$ galaxies and that satisfy the compactness criterion.

We call $n$ the number of accordant-velocity galaxies in the $mvCG$ and adopt $n_{\text{min}} = 4$ as our minimum number of accordant velocities.

Table 4 shows the number of groups in the observed and mock CG samples. The percentage of $mpCG$s that survive the velocity-filtering is 58 per cent (B06), 64 per cent (C06) and 60 per cent (DLB), so that our final samples of accordant velocity CGs contain from $\sim 2050$ to $\sim 2800$ $mvCG$s depending on the adopted SAM.

The main properties of the $mvCG$s identified in the DBL galaxy catalogue are shown in Fig. 2 together with the observed distribution of $vHCG$s (the distribution of the $mvCG$s obtained with the other two SAMs are similar, except for radial velocity distributions that are more skewed to lower values and considerably more groups in the bin of lowest group surface brightness). Hickson’s visual selection of CGs produced a catalogue that is incomplete at

#### Table 4. Compact group samples.

| Sample  | Type | gals | $v$ filter | $N$ | $\pi_{90}$ $(10^{-3} h^2 \text{Mpc}^{-3})$ |
|---------|------|------|------------|-----|---------------------------------|
| $pHCG$  | obs  | ext  | no         | 72  | 1.9                             |
| $vHCG$  | obs  | ext  | yes        | 52  | 1.1                             |
| $pmpCG$ | mock | par  | no         | B06 | C06 | DLB | 1.9 | 0.9 | 0.9 |
| $mpCG$  | mock | ext  | no         | C06 | DLB | 0.9 | 0.9 | 0.9 |
| $pmvCG$ | mock | par  | yes        | 4553 | 2685 | 5646 | 43 | 29 | 23 |
| $mvCG$  | mock | ext  | yes        | 2073 | 2095 | 2825 | 16 | 22 | 10 |
| $mvHC$  | mock | ext  | yes        | 272  | 223  | 291  | 5.1 | 5.4 | 2.6 |

Notes. Col. 1, sample; col. 2, sample type (obs, observed); col. 3, galaxy type (par, particle; ext, extended); col. 4, velocity filter; col(s). 5, number of groups (summed over eight vertexes for mocks); col(s). 6, space density within $90h^{-1}\text{Mpc}$ (eq. 1, divided by 8 for the mock samples to take into account the eight vertexes from which they were selected). Columns with three values show the results for the Bower et al. (2006) (B06), Croton et al. (C06), and De Lucia & Blaizot (DLB) SAMs, respectively. The limiting surface magnitude of the $mvHCG$ sample is not sharp.
Compact groups from the Millennium Simulations

1233

Figure 2. Distributions of properties of the velocity selected CGs: mock mvCGs from the De Lucia & Blaizot (2007) model (thin red histograms) and observed vHCGs (thick black dashed histograms). The properties are the group multiplicities (top left-hand panel), radial velocities (top right-hand panel), angular group diameters (middle left-hand panel), brightest galaxy magnitudes (middle right-hand panel), magnitude differences between the brightest galaxy and the full group (bottom left-hand panel) and group surface brightness (bottom right-hand panel). Error bars correspond to Poisson errors.

small angular sizes (middle left-hand plot), faint brightest galaxy magnitudes (middle right-hand plot), and in groups with a dominant brightest galaxy (bottom left-hand plot). We will quantify the completeness of the HCG in Section 2.9.

2.8 Testing the volume-limited sample

As we shall now see, limiting the depth of our galaxy sample to the simulation box provides a complete list of mvCG candidates. However, our neglect of galaxies further than the box size may prevent distant galaxies from spoiling the isolation of some of the mvCGs. Although the DLB model is available in an observing cone, this is not the case for the other two SAMs, and while we can construct a cone ourselves by placing galaxies of previous time-steps at the position corresponding to their look-back times, we do not have access to the $z > 0$ outputs of the Croton et al. (2006) model to do this, and we wish to consider all three SAMs in parallel.

We therefore use the DLB model to build a mock sample of galaxies within a cone, built of shells constructed from different snapshots corresponding to the epoch of the look-back time at their distance. Here we use the 17 last snapshots, bringing us to a maximum redshift of $z = 0.68$, where the minimum luminosity, $M_R = -24.91$, corresponds to $45 L_\odot$.

The top panel in Fig. 3 compares the radial velocity distribution of mpCGs extracted from the volume-limited mock catalogue described in Section 2.5 (solid lines) seen from one of its vertices with that obtained from groups identified in a magnitude-limited mock catalogue or light cone of the same solid angle (dashed lines). It can be clearly seen that the mpCGs obtained from the light cone and from the volume-limited catalogue before the $R_{\text{brightest}}$ cut-off (thin lines) are quite different. First, the cone sample is able to detect a large number of mpCGs beyond the limits of the box. On the other hand, fewer mpCGs are identified in the cone sample at small distances (up to 80 per cent of the box size). This lower abundance of mpCGs in the cone sample is the consequence of distant galaxies spoiling the isolation criterion of many mock compact groups. The total number of mpCGs in the cone sample is 4 per cent lower than in the box sample. These differences become more pronounced when the flux limit is applied (thick lines). The number of mpCGs in the cone sample is now 38 per cent lower than that of the box sample. Interestingly, the cone sample of mvCGs (with the brightest galaxy magnitude limit applied) shows a lack of CGs at all distances in comparison with the analogous box sample. The total number of mvCGs in the cone sample is half that of the analogous box sample. This indicates that our mpCG box-sample catalogue is only 62 per cent reliable against contamination of the isolation annulus by distant interlopers, while our mvCG box-sample catalogue is only 50 per cent reliable.

We can use this comparison of box and cone samples to correct the fraction of mpCGs that survive the velocity filter and make it as mvCGs: this fraction becomes $0.6 \times 0.5/0.62 = 48$ per cent $\pm 1$ (where the error is from binomial statistics and neglects the systematic error from the cone to box correction).
2.9 Completeness

2.9.1 Measure of completeness

It is interesting to compare the space density of mpCGs and mvCGs with those of the observed HCGs, selected in the same way. We estimate for the mock and observed samples the mean total surface density of CGs, as well as the mean space density of CGs within the lowest median distance of all samples, which is a fairly robust measure of density. The adopted distance is 9000 km s\(^{-1}\), which is close to the median of the vHCG sample, so the space density is computed as

\[
\pi_{90} = \frac{3 \int (v < 9000 \text{ km s}^{-1})}{90^3 \Delta \Omega} \text{ Mpc}^{-3},
\]

where \(v\) is the median velocity of the group members, while \(\Delta\) is the solid angle of the sample (in sr).

2.9.2 Projected compact groups

For the mpCGs, we obtain space densities \(\pi_{90} = 1.8, 2.3\) and \(1.1 \times 10^{-4} \text{ Mpc}^{-3}\) (see Table 4) using the SAMs by B06, C06 and DLB, respectively. In comparison, the HCGs were selected on the POSS I plates, spanning 9.7 sr = 32000 deg\(^2\) (Dec > -33\(^\circ\)). For the 72 mpHCGs, the mean density is \(\pi_{90} = 1.9 \times 10^{-5} \text{ Mpc}^{-3}\) (Table 4), i.e. typically nine times lower than the values obtained from the three samples of mpCGs.

Now, within a limiting distance of \(v = 9000\) km s\(^{-1}\), we found (Fig. 3) 17 mpCGs in our light cone in comparison with 22 in one of our boxes, again because our box sample misses possible distant interlopers that spoil the CG isolation. This suggests that we would have found 23 per cent fewer mpCGs, had we not limited ourselves to the box. We deduce that the observed pHCG sample is (1/19)/(17/22) \(\sim 14\) per cent complete at this limiting distance (which again corresponds to the median distance of the HCG catalogue).

Hickson’s inclusion of the Galactic plane should lead to under-estimates of the completeness of roughly 1/3, which is the fraction of his search area (\(\delta > -27^\circ\) covered by the POSS I survey) with low galactic latitudes \(|b| < 20^\circ\). Therefore, the bulk of the incompleteness of the HCGs lies in the incomplete visual selection at high galactic latitudes.

2.9.3 Velocity-filtered compact groups

We now compare the space density of mvCGs with that of the vHCG sample. For our mock samples of mvCGs with at least four accordant velocities, the space densities \(\pi_{90}\) are (Table 4) 1.6, 2.2 and \(1.0 \times 10^{-4} \text{ Mpc}^{-3}\), for B06, C06 and DLB, respectively. For comparison, for the 52 vHCGs (defined with at least four accordant velocities and with brightest galaxy magnitude brighter than 14.44), the space density is \(\pi_{90} = 1.1 \times 10^{-5} \text{ Mpc}^{-3}\) (Table 4). Therefore, the space density of mvCGs selected in the box is typically 15 times that of the observed vHCG.

However, within \(v < 9000\) km s\(^{-1}\), we found 16 mvCGs in our light cone versus 20 (20 per cent more) in our box (for a single vertex as observation point). This suggests that we would have found 20 per cent fewer mvCGs, had we not limited ourselves to the box (thus allowing for distant galaxies to spoil the isolation of these 20 per cent of the mvCGs). Therefore, we deduce that the completeness of the vHCG sample is 1/15/(16/20) = 8 per cent. Note that we assumed that the contamination of distant galaxies of the isolation criterion of mpCGs and mvCGs is independent of the SAM, even if we only measured this effect with the DLB model.

The top panels of Fig. 4 show the completeness of the velocity-filtered Hickson sample as a function of radial velocity for the three SAMs. The completeness is defined as \(C(v) = \pi_{\text{vHCG}} / \pi_{\text{vHCG}}\), where \(H_v = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}\). The green arrow shows the limit of our nearby subsample (see Table 4). The next five rows of panels show the completeness within the nearby subsamples (\(v < 9000\) km s\(^{-1}\)) \(C = \pi_{90} / \pi_{90}\), as a function of the other observable properties for the nearby subsample.

For all three SAMs, the vHCG completeness decreases sharply with distance (top panel), for groups with a dominant brightest galaxy (bottom panel), while the C06 model also predicts a decrease in vHCG completeness at fainter magnitude and lower surface brightness, whereas these trends are weaker in B06 and absent in DLB. Comparing the left and right sets of panels, one sees similar qualitative trends of completeness versus interesting parameter, but with overall completeness relative to the extended-galaxy mvCGs that is roughly double the value of the completeness relative to particle-based pmvCGs. Moreover, one can see a slightly faster trend of decreasing completeness with decreasing surface brightness with the C06 model for the mvCG in comparison to the pmvCGs.

2.9.4 Comparison with previous studies

Previous studies have concluded that the pHCG sample is incomplete at low group surface brightness (\(24 < \mu_R < 26\) mag arcsec\(^{-2}\)): Hickson (1982) and Walke & Mamon (1989) (from the lack of low surface brightness groups), and Prandoni et al. (1994) (from a comparison with their own automatically selected projected SCG sample of compact groups, built with very similar criteria as Hickson 1982). We also found a large incompleteness at low surface brightness when comparing the pHCGs with our mpCG samples. However, the sharp drop in the number of compact groups with low surface brightness (\(\mu_R > 24\) observed in the pHCGs, is also clearly visible in the mvCGs with the DLB model (but less so with the other two SAMs).

The incompleteness in brightest galaxy counts is analogous to the incompleteness in group number counts that Hickson had noticed at \(R = 13.0\) and that Prandoni et al. had already noticed at the much brighter limit of \(b_J = 13.1\) (from the break in the slope of the number counts away from the Euclidean value of 0.6).

Mamon (2000) noted that fig. 7 of Prandoni et al. indicates that the pHCG catalogue is incomplete by a factor of 3 at bright magnitudes, relative to the SCG catalogue, while this incompleteness gets worse at increasingly fainter magnitudes. A closer look at their fig. 7 reveals that the number of groups brighter than \(b_J = 13\) is roughly 30 for the HCG and 2.5 for the Euclidean extrapolation of the SCG group counts to this relatively bright magnitude. Given that the solid angle of the HCG (32000 deg\(^2\)) is 25 times that of the SCG (1300 deg\(^2\)), the completeness of the HCG relative to the SCG is 30/2.5/25 = 0.48, with total surface densities of 30/32000 = 0.9 \(\times 10^{-3}\) deg\(^{-2}\) and 2.5/1300 = 1.9 \(\times 10^{-3}\) deg\(^{-2}\) for the pHCG and SCG, respectively.

This strong incompleteness at faint magnitudes is also evident for vHCGs, as seen in the middle right-hand panel of our Fig. 2, which suggests (by matching the magnitude counts at intermediate magnitudes) that the HCG becomes incomplete for brightest galaxy magnitudes fainter than \(R = 12.5\), to the point where at \(R = 14.44\), the differential completeness falls to roughly 5 per cent. The surface
Figure 4. Completeness of the $vHCG$ catalogue relative to the $pmCG$ (left) and $mvCG$ (right) mock catalogues as a function of distance (top panels) and for groups within $v = 9000\ km\ s^{-1}$ (other five rows of panels). From top to bottom the panels represent global cumulative completeness versus distance, versus group total magnitude, versus group surface brightness and versus brightest group galaxy, and differential completeness versus angular size and versus dominance of brightest galaxy. The reference SAMs are Bower et al. (blue circles), Croton et al. (red triangles), and De Lucia & Blaizot (black crosses). Error bars for differential completeness are computed as binomial errors and are shown only for DLB SAM.

2.10 Mock Hickson compact groups

As noted above, the HCGs produced by Hickson’s visual inspection cannot be reproduced by an automatic searching algorithm given the many biases in the selection of HCGs. Therefore, the nature and properties of the mock CGs that strictly meet the HCG criteria mentioned in Section 2.6 may be different from the properties of the HCGs themselves.

Given the strong and progressive incompleteness of the HCGs in brightest galaxy counts, small angular sizes, and systems with strongly dominant brightest galaxies, it is essential to fold in these extra factors of incompleteness when building a sample that will be a good mock for the observed HCGs. We therefore wish to construct a mock velocity-selected Hickson compact group ($mvHCG$) sample, starting with the $mvCG$ sample, and selecting galaxies with probabilities proportional to the completeness in (1) group surface brightness, (2) brightest galaxy magnitude and (3) difference between the brightest galaxy and total group magnitude (i.e. the relative importance of the brightest galaxy). We do not consider the distribution of angular sizes, since this latter quantity is directly dependent on the three other parameters.

Because the resulting number of $mvHCG$s turns out to be very small, rather than select $mvCG$s according to the probability that a given $mvCG$ would be observed by Hickson, we proceeded as follows. We selected the first $mvCG$s that fill the observed distribution of $vHCG$s for the three parameters, and stopped once one of the 10 bins in any of the three distributions for the $mvCG$s reached the value observed in the corresponding bin for the $vHCG$s. Hence, the derived distributions of the three parameters do not match perfectly the observed ones, but are lower limits. We repeated this exercise, using different orders for our loop over the $mvCG$s until we matched as best as possible the observed $vHCG$ distributions.

This procedure was again applied on the eight different samples corresponding to the eight observers situated at the eight vertexes of the simulation cube obtaining final samples of typically 250 $mvHCG$s.

The distribution of properties of $mvHCG$s, shown in Fig. 5 for the DLB model, matches much better the observed distributions of $vHCG$s. Similar results are found using the other two SAMs. The three SAMs fail to find groups of roughly concordant magnitudes.

The three $mvHCG$ samples will be used to compare with the properties of observed $vHCG$s and with the correlations obtained for previous authors based on the observed accordant-velocity HCGs.

3 DIFFERENT CLASSES OF COMPACT GROUPS

Even though we have used redshift information to identify our $mvCG$s, the selected groups are not necessarily physically dense in 3D real space.
Our $mvCG$s can thus be split into three classes:

(i) physically dense groups (Real),
(ii) chance alignments within loose groups (CALG),
(iii) chance alignments within filaments (CAF).

Keeping with the original intent of Hickson (1982), who had selected in projection compact groups of at least four galaxies, we will classify an $mvCG$ as Real if at least four of its galaxies form a physically dense group. Also, we will sometimes join the CALG and CAF classes into the set of Chance Alignments (CAs).

There are several ways to use the three-dimensional information to define these classes, and as we shall see that none of them is perfect.

3.1 Binding energies

A simple way to separate the Real CGs from the CAs is to use the binding energy of the system. In Appendix B, we show that binding energies are highly inaccurate for groups of masses $M < 10^{14} h^{-1} M_{\odot}$, and cannot be used to distinguish which CGs are physically dense and which are caused by chance alignments.

3.2 Line-of-sight shape and 3D length

Alternatively, we can classify the mock CGs using their size and/or elongation. We consider the four closest galaxies in each mock CG, again in line with the original intent of Hickson (1982). By closest four galaxies we mean either the entire $mvCG$ if it has only four galaxies, or else the subgroup of four with the smallest 3D length. We use the following notations for these smallest quartets:

(i) $s$: maximum 3D separation, hereafter 3D length;
(ii) $S_{\perp}$: maximum projected separation, hereafter projected size;
(iii) $S_{\parallel}$: maximum line-of-sight separation, hereafter line-of-sight length;
(iv) $S_{\parallel}/S_{\perp}$: hereafter, line-of-sight elongation;
(v) round $mvCG$: $S_{\parallel}/S_{\perp} < 2$;
(vi) elongated $mvCG$: $S_{\parallel}/S_{\perp} \geq 2$.

Fig. 6 shows how the line-of-sight elongation is related to the maximum 3D separation. The points in the lower left-hand part of Fig. 6 show uncorrelated line-of-sight elongation and 3D length, as expected of Real groups, while the upper right-hand part of the figure shows instead a strong correlation of line-of-sight elongation with 3D length, indicative of CA groups. Which cuts in line-of-sight elongation and 3D length separate best the Real CGs from the CAs? The choice of the critical 3D length, $s_{\text{cut}}$, is not straightforward, as we shall now discuss.

3.2.1 Matching line-of-sight elongation of real-space selected groups

We first tried varying $s_{\text{cut}}$ by imposing that the median line-of-sight elongation be equal to that of real-space-selected groups. We measured a median line-of-sight elongation of 0.725 for the PG3Ds. We also checked this median value of the line-of-sight elongation with Monte Carlo simulations of quartets distributed at random in a virial sphere with an Navarro, Frenk & White (1996, hereafter NFW) density profile with concentration $r_{v}/r_{s} = 10$, where we then elongated the sphere in two orthogonal directions by two factors to make it a triaxial ellipsoid, and observed it from a random direction, and repeated this exercise 5000 times. We then find median line-of-sight elongations of 0.782 (in spheres) and 0.719 [in triaxial ellipsoids, with $b/a = 0.79$ and $c/a = 0.65$, as found by Jing...
Figure 7. Upper panels: Median line-of-sight elongation versus critical 3D length, $s_{\text{cut}}$, for the four galaxies in the richest subclump of mock velocity-accordant compact groups. Dashed (solid) lines refer to values of $s_{\text{cut}}$ where the sample of $mvCG$s has less (more) than 50 $mvCG$s. The red solid horizontal lines correspond to the median line-of-sight elongation measured by selecting four galaxies at random from each PG3D group. Lower panels: Normalized cumulative distribution of 3D lengths, considered as critical separations between the Real and CA classes. The green arrows indicate the value of $s_{\text{cut}}$ that matches the median line-of-sight elongations of the PG3D groups (vertical arrows) yielding the corresponding fraction of Real CGs for the adopted $s_{\text{cut}}$ (horizontal arrows).

Figure 8. Projected size versus line-of-sight length for the smallest quartets within $mvCG$s, using the De Lucia & Blaizot (2007) SAM. Also shown are log-spaced contours (red), the medians in bins of 200 points (green jagged line) the interquartiles (magenta jagged lines), and the line $y = x$ (thin blue line).

as round as real-space selected ones: redshift-space selection will always produce somewhat more elongated groups than real-space selected ones.

3.2.2 Line-of-sight elongation versus line-of-sight size

Alternatively, one could argue that CAs should be long in the absolute, i.e. high $s$, and/or relative to their projected sizes, i.e. high $S_{\parallel}/S_{\perp}$, e.g. $S_{\parallel}/S_{\perp} > 2$.

If the projected sizes were independent of the line-of-sight lengths, as would be expected if all CGs were CAs, we could then impose a value of $s_{\text{cut}}$ that would be close to $\sqrt{2}$ times the upper envelope of $S_{\parallel}$ (since round CGs would have $s \simeq \sqrt{S_{\parallel}^2 + S_{\perp}^2} \simeq \sqrt{2} S_{\parallel}$).

Fig. 8 shows that the $mvCG$s behave differently: while at a high line-of-sight length, where CAs are expected to be dominant, the projected size is indeed independent of the line-of-sight length, at a low line-of-sight length, where CAs are not dominant, the projected size increases with increasing line-of-sight length. So the upper envelope of the projected sizes is not a clear-cut value. The contours suggest a close to linear increase of $S_{\parallel}$ with $S_{\parallel}$ in the low $S_{\parallel}$ regime, as expected of systems of the same line-of-sight elongations and different sizes.

The transition between these two regimes is difficult to ascertain. One way is to look for the value of $S_{\parallel}$ for which the slope of the median $S_{\parallel}$ changes from steep to shallow. This yields a critical $S_{\parallel}$ of $\approx 140 \, h^{-1} \, \text{kpc}$ for the DLB model (see Fig. 8), $120 \, h^{-1} \, \text{kpc}$ for the B06 model and $165 \, h^{-1} \, \text{kpc}$ for the C06 model. These three critical values of $S_{\parallel}$ correspond to $S_{\parallel} \approx 100, 105$ and $112 \, h^{-1} \, \text{kpc}$, for the B06, C06 and DLB models, respectively. One therefore infers critical group length of $s_{\text{cut}} = 156, 196$ and $179 \, h^{-1} \, \text{kpc}$, for the B06, C06 and DLB models, respectively.

The fraction of $mvCG$s with lengths smaller than these three values of $s_{\text{cut}}$ can then be read from the bottom panel of Fig. 9: one finds that 72, 72 and 59 per cent of the groups have $s < s_{\text{cut}}$ for the B06, C06 and DLB models, respectively. One would therefore
deduce that between half and three-quarters of the \( mvCG \)s are Real (depending on the SAM). However, given the crudeness of the method, one should take these percentages with caution.

### 3.2.3 Reasonable cuts in length and line-of-sight elongation

We now explore whether reasonable limits on \( s_{\text{cut}} \) and \( S_\parallel /S_\perp \) can reduce substantially the fraction of Real \( mvCG \)s. Fig. 9 displays the distributions of the 3D length, \( s \), for the DLB model.

The distribution of 3D lengths clearly shows a dominant log-normal component and a second component of more extended lengths. In fact, when restricting to round groups \( (S_\parallel /S_\perp < 2 \text{, dashed histograms}) \), the round \( mvCG \)s tend to be smaller with the Bower model \( (s = 78 \, h^{-1} \text{kpc}) \) and larger with the DLB model \( (s = 125 \, h^{-1} \text{kpc}) \), with the predictions from the Croton model in between. Finally, the distribution of elongated \( mvCG \)s \( (S_\parallel /S_\perp \geq 2) \) is wider than that of the round \( mvCG \)s, and centred around \( s \simeq 250 \, h^{-1} \text{kpc} \) for the B06 and C06 models and 300 \( h^{-1} \text{kpc} \) for the DLB model. It displays an extended tail of very large \( (>2 \, h^{-1} \text{Mpc}) \) 3D lengths.

The discussion above suggests a conservative maximum for the Real \( mvCG \)s of \( s_{\text{cut}} < 200 \, h^{-1} \text{kpc} \).

### 3.3 Fraction of CAs in different samples of mock CGs

Table 5 summarizes the fractions of CGs satisfying various criteria that could classify them as Real. With our choice of \( s_{\text{cut}} = 200 \, h^{-1} \text{kpc} \), we obtain fractions of Real \( mvCG \)s of 0.80 (Bower), 0.73 (Croton) and 0.64 (DLB). To be more favourable to the CAs, we can include additional elongated \( mvCG \)s. However, it makes no sense to call a CA an elongated \( mvCG \) with a very small 3D length, for example with \( s = 50 \, h^{-1} \text{kpc} \), because such an \( mvCG \) is also a physically dense group, hence a Real. So, we consider a simple hybrid classification (dashed lines in Fig. 6), where the CAs are the \( mvCG \)s with \( s \geq 200 \, h^{-1} \text{kpc} \) OR \( (s > 100 \, h^{-1} \text{kpc} AND S_\parallel /S_\perp \geq 2) \). We then find (Table 5) that the fraction of Real \( mvCG \)s is 0.76 (Bower), 0.67 (Croton) and 0.59 (DLB). We hereafter adopt this hybrid criterion to estimate the fraction of Real and CA compact groups.

It therefore appears that more than half of the \( mvCG \)s are physically dense, although there are important variations between the three galaxy formation models, with Bower et al. predicting the most physically dense \( mvCG \)s, De Lucia & Blaizot predicting the least, and Croton et al. in between.

If we only consider CGs identified in projection (\( pmCG \)s), the Real CGs represent between 35 and 47 per cent, depending on the criteria and on the SAM (Table 5).

We note that the fraction of CAs diminishes in all three SAMs when going from the \( pmCG \) to \( mvCG \) and to \( HCG \) samples, i.e. when first taking into account the extended nature of galaxies causing confusion, and then in incorporating the biases we measured in the visual selection of Hickson (1982). Table 5 thus indicates that the fraction of CAs is 28–47 per cent for the \( pmCG \)s, but is reduced to 24–41 per cent for the \( mvCG \)s and only 14–30 per cent for the \( HCG \)s.

### Table 5. Fraction of real mock CGs using different criteria.

| Criterion | B06 | C06 | DLB |
|-----------|-----|-----|-----|
| \( pmCG \) | | | |
| \( s < 200 \, h^{-1} \text{kpc} \) | 0.46 | 0.39 | 0.22 |
| \( S_\parallel /S_\perp < 2 \) | 0.43 | 0.38 | 0.23 |
| \( s < 100 \, h^{-1} \text{kpc} \) OR \( (s < 200 \, h^{-1} \text{kpc} AND S_\parallel /S_\perp < 2) \) | 0.43 | 0.36 | 0.20 |
| \( mvCG \) | | | |
| \( s < 200 \, h^{-1} \text{kpc} \) | 0.46 | 0.47 | 0.38 |
| \( S_\parallel /S_\perp < 2 \) | 0.46 | 0.46 | 0.43 |
| \( s < 100 \, h^{-1} \text{kpc} \) OR \( (s < 200 \, h^{-1} \text{kpc} AND S_\parallel /S_\perp < 2) \) | 0.44 | 0.43 | 0.35 |
| \( pmvCG \) | | | |
| \( s < 200 \, h^{-1} \text{kpc} \) | 0.77 | 0.70 | 0.59 |
| \( S_\parallel /S_\perp < 2 \) | 0.71 | 0.68 | 0.61 |
| \( s < 100 \, h^{-1} \text{kpc} \) OR \( (s < 200 \, h^{-1} \text{kpc} AND S_\parallel /S_\perp < 2) \) | 0.72 | 0.64 | 0.53 |
| \( mvHCG \) | | | |
| \( s < 200 \, h^{-1} \text{kpc} \) | 0.80 | 0.73 | 0.64 |
| \( S_\parallel /S_\perp < 2 \) | 0.80 | 0.72 | 0.71 |
| \( s < 100 \, h^{-1} \text{kpc} \) OR \( (s < 200 \, h^{-1} \text{kpc} AND S_\parallel /S_\perp < 2) \) | 0.76 | 0.67 | 0.59 |

© 2010 The Authors. Journal compilation © 2010 RAS, MNRAS 409, 1227–1243
3.4 Chance alignments within loose groups and beyond

We now consider as CAs all $mvCG$s with 3D lengths $s \geq 200h^{-1}$ kpc, regardless of the line-of-sight elongation. The distinction between CALG and CAF is very simple, as we simply check whether all CALG members lie within a single PG3D (making it a CALG) or not (making it a CAF).

Table 6 shows that large majority of the CAs (typically three-quarters) are CALGs, regardless of the SAM and the sample. Thus, alignments within filaments or with galaxies in the field are much less likely than chance alignments within larger groups or than having physically dense groups.

3.5 Isolated dense groups?

The standard picture is that dense groups of galaxies are the cores of virialized looser groups. However, some HCGs appear extremely isolated (Rood & Williams 1989; Palumbo et al. 1995). The simulations that we have analysed allow us to check whether dense groups can be isolated out to the virial radius. We have simply cross-identified the $mvCG$s with the real-space-selected PG3Ds. We then call a CG isolated if it constitutes the entire PG3D (in the magnitude range determined by the brightest galaxy of the CG). We then find that only $\sim$11 per cent of the $mvCG$s constitute the entire PG3D in all the three SAMs. In the $mvHCG$ samples, the fraction of isolated CGs are somewhat smaller (7 per cent for B06, 6 per cent for C06 and 8 per cent for DLB).

4 DISCUSSION AND CONCLUSIONS

4.1 Summary of results

The aim of this paper is twofold: (1) predicting the nature of automatically identified CGs, and (2) predicting the nature of the well-studied but highly incomplete and biased HCGs. We identify CGs in three mock galaxy catalogues, constructed from the Millennium Simulation at $z = 0$ combined with three semi-analytical models (Bower et al. 2006; Croton et al. 2006; De Lucia & Blaizot 2007). Several thousand mock CGs are identified using a two-dimensional automated algorithm similar to that applied by Hickson (1982) plus a restriction in the brightest galaxy magnitude. We also allowed for CGs that contain isolated and compact subgroups and furthermore considered the important effects of extended galaxies causing confusion for close projected pairs. Our main results are as follows.

(i) Among the observable $mpCG$s, $\sim$60 per cent have at least four galaxies within 1000 km s$^{-1}$ of the median velocity of the group, regardless of the SAM. We tested whether our results might be biased by the redshift cut-off of our sample ($z \sim 0.17$). We find that identifying in our (magnitude+)$v$olume limited catalogue produces 1.6 times more $mvCG$s than identifying on an only-magnitude-limited catalogue, because our box sample misses distant galaxies that spoil the isolation of mock compact groups. Correcting for this effect, we deduce that the fraction of $mvCG$s that survive the velocity filter is 50 per cent. In comparison, there are as many as 52 $vHCG$s among 72 $pHCG$s, meaning that 71 per cent of observed HCGs survive the velocity filter. From binomial statistics, the probability that we find as many as 52 $vHCG$s given that 36 (half of 72) are expected is negligible ($P = 10^{-4}$).

(ii) Comparing the space densities of the $mvCG$s and the $vHCG$s, we deduce that the HCG catalogue is only 8 per cent complete. A comparison of the parameter distributions between the $mvCG$s and $vHCG$s indicates that the HCG is incomplete in groups of small angular sizes, high fraction of light in the brightest galaxy, as well as (for the C06 model) faint brightest galaxy magnitudes and low surface brightness.

(iii) We find that the velocity filtering of $mpCG$s does not necessarily imply that the resulting accordant–velocity CGs are physically dense. We tested different criteria to classify the accordant–velocity CGs according to the maximum 3D galaxy separations and the line-of-sight elongations. We find that, with the most conservative criterion, at least 3/5 of the mock accordant–velocity CGs are physically dense, although the precise fraction depends on the galaxy formation model used.

(iv) The large majority of non-Real mock accordant–velocity CGs are caused by chance alignments within larger groups, rather than within larger regions such as large-scale filaments.

(v) We find that the fraction of chance alignments decreases from 28–47 per cent for the particle-based $mvCG$s to 24–41 per cent for the sample with close pairs removed ($mvCG$s) to only 14–30 per cent once we fold in the biases of the HCG ($mvHCG$s). This explains in part why simulation studies (Mamon 1986; Walk & Mamon 1989; Hernquist et al. 1995) predict more chance alignments than one infers from actual observations.

Table 4 indicates that typically half of the groups are lost once we apply the criterion to blend close projected pairs of galaxies. While the discarded groups no longer satisfy the selection threshold of four galaxies within 3 mag from the brightest, there are still many blended pairs within the groups that survived the blending criterion (in 1/3, 1/4 and 1/8 of the mock CGs for B06, DLB and C06, respectively, with similar fractions for $mpCG$s and $mvCG$s). Could this mean that a significant fraction of the $PHCG$s and $vHCG$s contain blended galaxies?

A comparison of the lists of Hickson (1982) (selected from photographic plates) and Hickson et al. (1989) (CCD-based) indicates that the latter found 10 HCGs with extra galaxies (24, 26 $\pm$ 3, 27, 43, 51, 70, 72, 76 $\pm$ 2, 83 and 99), plus one with one galaxy less (HCG 40). So 10 per cent of the original HCG groups contained blended galaxies as discovered with the better CCD photometry. It therefore appears that our fractions of 1/8 to 1/3 of $mpCG$s with blended pairs is high. But if the truth is in between our particle case and our extended case, then we should have of order 6–17 per cent of groups with blended galaxies, which is in rough agreement with what we see in the HCGs.

Now, three HCGs have been observed at much higher resolution with HST imaging. One shows no extra galaxies (HCG 87), while the other two show more interacting units than counted by Hickson et al. (1989): HCG 31 seems to have seven galaxies and not just four, while HCG 90 has five galaxies and not four. Binomial statistics suggest that, with 95 per cent confidence, the view of two groups out of three with extra galaxies implies that the fraction of such groups with blended pairs is between 14 and 86 per cent. But HST is probably biased towards observing dense interacting HCGs. Still,
there could be interacting pairs showing galaxies that have been
blended even with the CCD images of Hickson et al. (1989). So,
with the high resolution of the HST, the fraction of pHCGs and
vHCGs with blended pairs may be considerably higher than 10 per
cent and in agreement with the fractions found in the mock CGs.

We can also estimate the number of HCGs that are physically
dense groups of at least four galaxies. As discussed in Section 2.1,
the HCG catalogue has 100 members, among which 99 are compact
groups (since ICG 54 is a collection of H II regions), of which only
83 actually fulfil the original magnitude concordance criterion (R-
band magnitude range less than 3). Among the 72 pHCGs whose
brightest and faintest galaxies are brighter than $R = 14.44$ and $R =
17.44$, respectively, only 52 (72 per cent) have at least four accordant
velocities, and among these, we expect roughly between 36 and
44 HCGs that are physically dense groups of at least four galaxies.
Extrapolating to the 68 accordant–velocity HCGs (including those
with magnitude range greater than 3 mag), we expect no more than
$\sim 58$ physically dense HCGs with at least four galaxies. In
comparison, Mamon (1986) had predicted that 47 out of what
he thought would be 78 accordant velocity HCGs are caused by
chance alignments (60 per cent), while the remaining 40 per cent
are physically dense (but he predicted that half of these dense groups
were unbound systems). We are therefore less pessimistic than
Mamon (1986) on the fraction of chance alignments polluting the
HCG catalogue, since chance alignments appear to represent be-
tween 14 and 30 per cent of the $m$vHCGs (we were not able to
check what fraction of the Real ones are unbound: see Appendix
B). Part of this discrepancy is caused by Mamon’s (1986) reliance
on simulations without consideration of selection effects such as
observers blending close projected pairs of galaxies. Still, the per-
centage of chance alignments in the particle mock velocity-filtered
compact groups ($p$mCGs) is only 28–47 per cent (depending on
the SAM). Nevertheless, given the wide range of chance alignments
fractions among the three SAMs, one cannot rule out that a more
realistic galaxy formation model would lead to as much as 60 per
cent of chance alignments.

4.2 Comparison with McConnachie et al.

McConnachie et al. (2008) (MEP) have published a study very
similar to ours: they also extracted $p$mCGs from the DLB model
obtained from the Millennium dark matter simulations. Their sam-
ple extended to $r = 18$, which corresponds to roughly one-quarter
of a magnitude fainter than our limit of $R < 17.44$. Their other $p$mCG
criteria appear to be almost exactly the same as ours (following the
criteria of Hickson 1982), although their algorithm works differ-
ently (McConnachie et al. 2009). MEP found a total of over 15 000
$p$mCGs over $4\pi$ sr.

Using precisely the same input galaxy catalogue (from Blaizot
et al.) as MEP, we find 25 000 $p$mCGs, so, our algorithm is nearly
1.6 times more efficient than MEP’s in finding $p$mCGs. Surpris-
ingly, if we build a light cone as we did in Section 2.8, with apparent
magnitude limit of $R = 17.67$ ($\approx 5g$ mag), we obtain a mock
galaxy catalogue that is three times denser than the Blaizot et al.
mock galaxy catalogue used by MEP. From our mock galaxy cata-
logue, we extract 15 191 $p$mCGs in 1.2693 sr, which means that,
with the data used in this work, we are $\sim 10$ times more efficient
than MEP in finding CGs, principally by differences in the parent
samples of galaxies, but also by differences in the CG detection
algorithm.

MEP and us agree that a significant fraction of $m$vCGs are caused
by chance alignments: MEP found 71 per cent of their $p$mCGs are
CAs while we find 80 per cent (with our hybrid classification, see
Table 5).

There are, however, several important differences between our
study and MEP’s study.

(i) MEP build their sample from a mock that extends beyond the
box size of the Millennium simulation — using the output of the
Mock Map Facility (MoMaF) code of Blaizot et al. (2005), while
our mock galaxy catalogues are limited to the size of the simulation
box. However, as shown in Section 2.8, working on a light cone
or working on a single simulation box leads to similar numbers of
mock CGs (we identify a factor of 1.2 fewer mpCGs, and 1.5 more
mvCGs).

(ii) MEP consider CGs with a faint magnitude limit, while we
also tie in a bright magnitude limit to ensure that all mock CGs were
built from galaxies that spanned a range of over 3 mag.

(iii) We have analysed the galaxies from three different SAMs,
while MEP have only considered the DLB sample, which we found
to produce the smallest fraction of physically dense mpCGs.

(iv) MEP only provide statistics for the mock CGs defined in
projection ($mpCGs$, which they refer to as ‘HA’s) but do not con-
sider the subset of accordant–velocity groups ($pvCGs$). We think
this would have been worthwhile because ever since Hickson et al.
(1992) published the HCG galaxy redshifts, most analyses have
thrown out the discordant velocity HCGs.

(v) MEP did not consider selection effects, while we considered
both the galaxy confusion from close, blended, projected pairs,
as well as the biases that we determined for the Hickson’s visual
selection of the HCGs.

(vi) MEP only considered those mpCGs with $k \geq 3$ galaxies
that lie very close in real space, while we considered $k \geq 4$ (to be
consistent with Hickson’s initial motivation to have at least four
galaxies per HCG).

(vii) MEP define the Real mpCGs using a Friends-of-Friends
linking length in real space, while we use a maximum real-space
separation and the elongation along the line of sight. Structures
built from small numbers of components with Friends-of-Friends
algorithms tend to be more filamentary (e.g. Moore, Frenk & White
1993). For mpCGs that are CALGs or CAFs, the most distant outlier
will determine a similar maximum length and critical linking length.
However, for mpCGs without outliers (e.g. Real mpCGs), the linking
length will be smaller than the 3D length. In other words, selecting
Real groups with a linking length of $200 h^{-1}$ kpc will result in group
3D lengths considerably greater. Moreover, for mpCGs with both
foreground and background galaxies, MEP’s $\ell$ must be compared
to our $half$-maximum size $s_{\text{real}}$/2, and here is there a discrepancy of
a factor of 2. Worse, for those (admittedly rare) cases of, say four,
galaxies aligned along the line of sight at roughly equal separations
just below $\ell$, one will end up with a group that spans up to $3 \ell =
600 h^{-1}$ kpc, which is now three times our maximum 3D length, but
will still be called Compact Association (Real) by MEP, although
it clearly is a chance alignment. In summary, this point and the
previous one imply that MEP’s criterion for calling an mpCG Real
is much more liberal than ours.

4.3 Perspectives

In forthcoming papers, we will analyse the distribution of and cor-
relations between the physical characteristics of the $mv$CGs, and
show how they depend on their classification, in view of optimiz-
ing the probability that a CG selected in redshift space is physi-
cally dense. It would be worthwhile to probe the formation of the
phatically dense CGs by analysing the merger trees of galaxies in the mock CGs. Finally, the analysis presented here will need to be confirmed with increasingly realistic simulations of galaxy catalogues, for example those constructed from future galaxy formation models run on the recent high-resolution Millennium-II dark matter simulation (Boylan-Kolchin et al. 2009), and also on future high-resolution cosmological hydrodynamical simulations, with realistic prescriptions for feedback from AGN and supernovae.

ACKNOWLEDGMENTS

We thank Paul Hickson, Abilio Mateus and Jack Sulentic for useful exchanges and an anonymous referee for interesting comments. The Millennium Simulation data bases used in this paper and the web application providing online access to them were constructed as part of the activities of the German Astrophysical Virtual Observatory. We thank Richard Bower, Darren Croton, and Gabriele De Lucia for allowing public access for the outputs of their very impressive semi-analytical models of galaxy formation. This research has made use of VizieR data base maintained by the Center de Données Stellaires in Strasbourg, France, and the NASA/IPAC Extragalactic Data base (NED) which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. This work was partially supported by the European Commission’s ALFA-II programme through its funding of the Latin-American Virtual Network for Astrophysics and Cosmology (LENAC), Consejo de Investigaciones Científicas y Técnicas de la República Argentina (CONICET), the Agencia Nacional de Promoción Científica y la Secretaría de Ciencia y Técnica de la Universidad Nacional de Córdoba (SeCyT).

REFERENCES

Allam S. M., Tucker D. L., 2000, Astron. Nachr., 321, 101
Arkhipova V. P., Afanasev V. L., Dostal V. A., Zasov A. V., Karachentsev I. D., Noskova R. I., Saveleva M. V., 1981, SvA, 25, 277
Bames J., 1985, MNRAS, 215, 517
Barton E., Geller M., Ramella M., Marzke R. O., da Costa L. N., 1996, AJ, 112, 871
Barton E. J., de Carvalho R. R., Geller M. J., 1998, AJ, 116, 1573
Bertone S., De Lucia G., Thomas P., 2007, MNRAS, 379, 1143
Blair J., Wadadekar Y., Guiderdoni B., Colombi S. T., Bertin E., Bouchet F. R., Devriendt J. E. G., Hatton S., 2005, MNRAS, 360, 159
Bode P. W., Cohn H. N., Luger P. M., 1993, ApJ, 416, 17
Bower G. R., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150
Bryan G. L., Norman M. L., 2002, ApJ, 577, 108
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245
Cavaliere A., Santangelo P., 1981, ApJ, 249, 449
Diaferio A., Geller M. J., Huchra J., 1994, AJ, 107, 1623
Diaferio A., Geller M. J., Ramella M., 1994, AJ, 107, 868
Dos Santos M., Mamon G. A., 1999, A&A, 352, 1
Eke V. R., Baugh C. M., Cole S., Frenk C. S., Navarro J. F., 2006, MNRAS, 370, 1147
Falco E. E. et al., 1999, PASP, 111, 438
Focardi P., Kelm B., 2002, A&A, 391, 35
Fukugita M., Shimasaku K., Ichikawa T., 1995, PASP, 107, 945
Hernquist L., Katz N., Weinberg D. H., 1995, ApJ, 442, 57
Hickson P., 1982, ApJ, 255, 382
Hickson P., 1993, Astrophys. Lett. Commun., 29, 1
Hickson P., Rood H. J., 1988, ApJ, 331, 1651
Hickson P., Ninkov Z., Huchra J., Mamon G., 1984, in Mardirossian F., Giuricin G., Mezzetti M., eds, Clusters and Groups of Galaxies. Reidel, Dordrecht, p. 367
Hickson P., Kindl E., Huchra J. P., 1988, ApJ, 331, 64
Hickson P., Kindl E., Aumann J. R., 1989, ApJS, 7, 687
Hickson P., Mendes de Oliveira C., Huchra J. P., Palumbo G. G., 1992, ApJ, 399, 353
Iovino A., 2002, AJ, 124, 2471
Iovino A., de Carvalho R., Gal R., Odewahn S., Lopes P., Mahabal A., Djorgovski S., 2003, AJ, 125, 1660
Jiang C. Y., Jing Y. P., Faltenbacher A., Lin W. P., Li C., 2008, ApJ, 675, 1095
Jing Y. P., Suto Y., 2002, ApJ, 574, 538
Kelm B., Focardi P., 2004, A&A, 418, 937
Lee B. C. et al., 2004, AJ, 127, 1811
Lokas E. L., Mamon G. A., 2001, MNRAS, 321, 155
Mamon G. A., 1986, ApJ, 307, 426
Mamon G. A., 1987, ApJ, 321, 622
Mamon G. A., 1989, A&A, 219, 98
Mamon G. A., 1992, in Mamon G. A., Gerbal D., eds, Distribution of Matter in the Universe, DAEC mtg. No. 2, Obs. de Paris, Paris, p. 51 (ftp://ftp.iap.fr/pub/from_users/gam/PAPERS/daec92cg.ps.gz)
Mamon G. A., 2000, in Valtonen M. J., Flynn C., eds, IAU Coll. No. 174, Small Galaxy Groups, Vol. 209, Astron. Soc. Pac., San Francisco, p. 217
Mamon G. A., 2008, A&A, 486, 113
Mateus A. 2008, MNRAS, submitted (arXiv:0802.2720)
Mateus A., Jimenez R., Gaztañaga E., 2008, ApJ, 684, L61
McConnachie A. W., Ellison S. L., Patton D. R., 2008, MNRAS, 387, 1281 (MEP)
McConnachie A. W., Patton D. R., Ellison S. L., Simard L., 2009, MNRAS, 395, 255
Mei S. et al., 2007, ApJ, 655, 144
Mendes de Oliveira C., Hickson P., 1991, ApJ, 380, 30
Mendes de Oliveira C. L., Temporin S., Cyriano E. S., Plana H., Amram P., Sodré L. J., Balkowski C., 2006, AJ, 132, 570
Moles M., del Olmo A., Perea J., Maregosa J., Marquez I., Costa V., 1993, A&A, 285, 404
Moore B., Frenk C. S., White S. D. M., 1993, MNRAS, 261, 827
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563 (NFW)
Ochsenbein F., Bauer P., Marcout J., 2000, A&AS, 143, 23
Ostriker J. P., Larson R. B., Ostriker E. F., 1983, ApJ, 272, 401
Poggianti B. M., 1997, A&A, 322, 399
Porquet D., 2004, ApJ, 603, 621
Prada F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563 (NFW)
Renaud S., Sanchez-Conde M. A., 2006, ApJ, 645, 1001
R serious for feedback from AGN and supernovae.
Rubin V. C., Hunter D. A., Ford W., Kent J., 1990, ApJ, 365, 86
Salpeter E. E., 1955, ApJ, 121, 161
Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, MNRAS, 343, 978
Springel V. et al., 2005, Nat, 435, 629
Stephan M. E., 1877, CR Acad. Sci., 84, 641
Sulentic J. W., 1987, ApJ, 322, 605
Sulentic J. W., 1997, ApJ, 482, 640
Tovmassian H. M., Yam O., Tiersch H., 2001, Rev. Mex. Astron. Astrof., 37, 173
Tovmassian H., Plionis M., Torres-Papaqui J. P., 2006, A&A, 456, 839
Verdes-Montenegro L., Del Olmo A., Iglesias-Páramo J. J., Perea J., Vilchez J. M., Yun M. S., Huchneimer W. K., 2002, A&A, 396, 815
Walké D. G., Plionis M., 1989, A&A, 225, 291
Xu G., 1995, ApJS, 98, 355
Zepf S. E., 1993, ApJ, 407, 448

APPENDIX A: FROM R-SDSS TO R-JOHNSON APPARENT MAGNITUDE LIMIT

According to table 3 of Fukugita, Shimasaku & Ichikawa (1995), \( r - R \) has typical values of 0.36 (E), 0.31 (S0), 0.33 (Sab), 0.32 (Sbc), 0.30 (Scd) and 0.20 (Im). We thus restrict the HCG sample to a total extrapolated R-band extinction-corrected magnitude

\[
R'_g = B'_g - (B - R)_{iso} + (B_T - B'_T) \left( 1 - \frac{A_B/A_V}{A_g/A_V} \right) < 17.44, \tag{A1}
\]

where we adopted \( A_B/A_V = 1.33 \) and \( A_g/A_V = 0.75 \) (Cardelli, Clayton & Mathis 1989). Equation (A1) assumes no \( B - R \) colour gradient in the galaxy.

APPENDIX B: GROUP BINDING ENERGIES

The binding energy of the particles of a mock CG is difficult to determine because it is not clear which particles belong to the CG. One could alternatively use the galaxies instead of the particles. But most of the mass of a group is thought to lie in between the galaxies, so one must be careful on how the binding energy analysis is performed. We separate the group into the system of galaxies and the remaining intergalactic dark matter. The kinetic energy of the group would then be

\[
K = K_g + K_d = \frac{3}{2} \left( M_g + M_d \right) \sigma_v^2, \tag{B1}
\]

where \( M_g \) and \( M_d \) are the masses of the galaxies and of the intergalactic dark matter, respectively, while \( \sigma_v \) is the one-dimensional velocity dispersion assumed to be the same for the intergalactic dark matter and the galaxies, and where we placed ourselves in the group centre of mass to get rid of the bulk kinetic energy. The potential energy is more difficult to handle as it is the sum of the potential energies of the galaxy system, the intergalactic dark matter system and the cross-term between galaxies and the intergalactic dark matter. Suppose a group has a factor of \( m \) more intergalactic dark matter than mass in galaxies. From equation (B1), the kinetic energy is then

\[
K = (\mu + 1) K_g, \tag{B2}
\]

where \( \mu = \frac{m}{m + 1} \), hence to a virial ratio of the galaxies in a group can be far off from unity! If the intergalactic dark matter makes up for say \( \mu = 4 \) times as much as the galaxy mass, then if the group is in virial equilibrium, equation (B5) leads to \( -2 \frac{K_{gg}}{W_{gg}} = 21/5 = 4.2 \). Therefore, \( \mu = 0 \), hence to a virial ratio of unity according to equation (B5).

The simulation data that we have at our disposal provide the virial masses of the PG3D groups (dark matter included). If a mock CG is a CA, then we cannot know how much dark matter is assigned to this CA, but only to the PG3D group associated with it. We therefore choose to scale the galaxy masses to the PG3D group mass, i.e.

\[
m' = \frac{m}{\sum_{i=PG3D} m_i} M_{PG3D}. \tag{B6}
\]

Still, there remains the issue of CG galaxies that do not belong to any PG3D. One possibility is to apply equation (B6) to the CG galaxies that lie within PG3D groups, without scaling the masses of the isolated galaxies. The alternative is to scale by the fraction of mass in the whole simulation box, instead of the PG3D group.

Another issue is that the virial ratio of unity according to equation (B5). One way to avoid these problems is to assign to each galaxy the fraction of the total group mass equal to the ratio of its mass divided by the total mass in galaxies for that group. In other words, we are putting the intergalactic dark matter mass in each galaxy in proportion to its mass. This is equivalent to \( \mu = 0 \), hence to a virial ratio of unity according to equation (B5).

The virial theorem becomes

\[
2 K + \sum_i F \cdot r = 0, \tag{B8}
\]

where the Clausius virial of a group is

\[
\sum_i F \cdot r = \sum_i r_i \cdot \nabla \ln V_{int} (r_{ij}) = \sum_i \frac{G m_i m'_j}{r_{ij}^2 + r_{ij, rms}^2} \frac{r_{ij}^2}{r_{ij}^2 + r_{ij, rms}^2} \tag{B9}
\]

where \( m_j \) is the one-dimensional velocity dispersion of the galaxy. One way to avoid these problems is to assign to each galaxy the fraction of the total group mass equal to the ratio of its mass divided by the total mass in galaxies for that group. In other words, we are putting the intergalactic dark matter mass in each galaxy in proportion to its mass. This is equivalent to \( \mu = 0 \), hence to a virial ratio of unity according to equation (B5).
But how do we estimate the galaxy half-mass radii? We can compute analytically the half-mass radius of the matter within the virial radius, say for an NFW model, with a concentration \( c = r_v/r_{-2} = 10 \) (where \( r_{-2} \) is the ‘scale’ radius of slope \(-2\)), for which \( r_{h}/r_v \simeq 0.36 \) (Łokas & Mamon 2001, fig. 4 and equation 28). But we could also compute the half-mass radius within a larger radius, say the turnaround radius beyond which the Universe is expanding, and which is typically 3.5 times the virial radius. Assuming that the NFW model extends that far (see Prada et al. 2006), going to the turnaround radius amounts to increasing the concentration by a factor of \( r_{ta}/r_v \approx 3.5 \). So, if \( c = 10 \) for a galaxy, at the turnaround radius, we would use \( c = 35 \) and find \( r_{h}/r_{ta} \simeq 0.23 \), i.e. \( r_{h}/r_v \simeq 0.79 \).

But then, how do we estimate the mass within the virial radius of the galaxy? We could guess a mass-to-light ratio \( M(r_v)/L_B = 100 \), although \( M(r_v)/L_B \) is thought to decrease with increasing luminosity to reach a minimum around 70 (Eke et al. 2006). Then \( r_v = \left[ 2/\Delta \left( GM/L\right)^2 \right]^{1/3} = \left[ 2/\Delta \left( G(M/L)H_0^2 \right) \right]^{1/3} = 544 (L/10^{11})^{1/3} \) kpc for \( \Delta = 100 \) (as we used in the paper), \( H_0 = 73 \) km s\(^{-1}\) Mpc\(^{-1}\) (as used in the Millennium Simulation) and \( M/L = 100 \). For \( L_* = 0.18 \times 10^{11} L_\odot \), we end up with \( r_v = 307 \) kpc. So for \( L_* \) galaxies, we need a softening of typically \( r_0 = 70 \) to 250 kpc, i.e. \( c_0 = r_0/L^{1/3} = 0.03 \) to 0.10. Of course, the higher the softening scale, the less negative is the potential energy and the less bound is the system.

We test our prescription by computing the virial ratios of the PG3Ds using both PG3D scaling and box scaling of the galaxy masses, with different values for the softening scale \( c_0 \). The correct scaling must lead to virial ratios of unity, independent of group mass.

Fig. B1 shows the results of our test on PG3Ds. For both normalizations, we find that the softening \( c_0 = 0.03 \) kpc \( L_\odot \)^{1/3} (corresponding to \( r_0 = 70 \) h\(^{-1}\) kpc for \( L = L_* \) galaxies) bring the virial ratios of the highest-mass PG3Ds to unity. Without the softening, the potential energies are overestimated (in absolute value), hence the virial ratios are underestimated, while with too strong softening the virial ratios are overestimated.

However, Fig. B1 indicates that even with the correct softening, i.e. with correct virial ratios at the high-mass end, there is so much scatter in the virial ratios at low masses, that over 13 per cent of the PG3Ds are found to be unbound for PG3D virial masses below \( 4 \times 10^{13} M_\odot \). This means that for the typical masses of the CGs, our virial ratio estimator is too inaccurate to use as a CG classifier.