

# The Rotational Velocity Distribution Function of Massive Disc Galaxies

by  
Juan José Fernández Morales

And tutored by  
Dr. Claudio Dalla Vecchia

Trabajo de Fin de Grado

Universidad de La Laguna  
San Cristóbal de La Laguna, Tenerife  
July 7, 2021

# Abstract

Español

*El objetivo del proyecto es caracterizar la función de distribución de velocidad de rotación de galaxias de disco masivas mediante el uso de simulaciones cosmológicas e hidrodinámicas de última generación de EAGLE de formación y evolución de galaxias. La función de distribución se comparará con las observaciones y, finalmente, se correlacionará con las propiedades de las galaxias y el entorno en el que se formaron las galaxias.*

**Palabras clave:**  $\Lambda$ CDM, EAGLE, simulación, velocidad máxima de rotación, masa estelar

English

*The aim of the project is to characterise the rotational velocity distribution function of massive disc galaxies by using the state-of-the-art EAGLE cosmological, hydrodynamic simulations of galaxy formation and evolution. The distribution function will be compared with observations, and eventually correlated to properties of galaxies and the environment in which galaxies formed.*

**Keywords:**  $\Lambda$ CDM, EAGLE, simulation, maximum rotational velocity, stellar mass

# Acknowledgments

*Un especial agradecimiento a todo profesor que me dejó preguntar y me quiso responder. Gracias a Claudio por querer formar parte de este proyecto. A mi familia, porque siempre les siento cerca a pesar de la distancia. Y a los amigos; la cantidad de motivos es tal y tan variada que si los enumero superan la página de agradecimientos, siempre les llevo conmigo.*

*L ywjw lpria pri zi jidw ikyeulu zek lalk.  
L awk pri aik yrvi pri jibeu ljewk,  
L awk pri aik qrji jibeu fwal.*

*Para Johanna, Victor e Himar*

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgments</b>	<b>ii</b>
<b>1 Introduction to EAGLE</b>	<b>1</b>
1.1 A brief visit to $\Lambda$ CDM . . . . .	1
1.2 The Importance of Numerical Simulations. . . . .	4
1.3 Inside EAGLE . . . . .	5
1.3.1 Initial Conditions and Cosmological Parameters . . . . .	6
1.3.2 Snapshots . . . . .	7
1.3.3 Friends-of-Friends Algorithm . . . . .	7
1.3.4 How to Extract the Information . . . . .	7
<b>2 Objectives and Methodology</b>	<b>9</b>
2.1 Morphology of Disc Galaxies . . . . .	10
2.2 Rotational Velocity Distribution Function . . . . .	11
2.2.1 Tully-Fisher Relation . . . . .	12
2.2.2 Distribution Function . . . . .	13
2.3 Properties under Initial Conditions. . . . .	13
2.3.1 Internship Code Modifications . . . . .	14
2.3.2 TFG Code Algorithm. . . . .	16
2.3.3 Over-density and Angular Momentum . . . . .	20
<b>3 Discussion of the Results and Conclusion</b>	<b>23</b>
3.1 Ref-L0050N0752 Simulation . . . . .	23
3.2 Tully-Fisher Relation . . . . .	24
3.3 Distribution Functions . . . . .	25
3.4 Initial Conditions . . . . .	27
3.4.1 The New Code . . . . .	27
3.4.2 Over-density . . . . .	27
3.4.3 Angular Momentum . . . . .	30
3.5 Conclusions . . . . .	31
<b>A Code to Obtain the Dataset</b>	<b>36</b>

# List of Figures

2.1	$k_{co}$ versus $M_*$ for massive galaxies ( $M_* > 10^9 M_\odot$ ) in Ref-L0050N0752 at $z = 0$ (snapshot 28)	10
2.2	Different perspective of a massive disc galaxy of Ref-L0050N752 at $z = 0$ (snapshot 28)	11
2.3	$V_{\max}$ versus $M_*$ for massive galaxies ( $M_* > 10^9 M_\odot$ ) in Ref-L0050N0752 at $z = 0$ (snapshot 28)	12
2.4	$V_{\max}$ and $M_*$ distributions for massive galaxies in Ref-L0050N0752 at $z = 0$ (snapshot 28)	15
2.5	Comparison of galaxy stellar mass functions.	15
2.6	Representations of some of the phases of the written code during the practice period	17
2.7	Over-density module histogram for massive disc galaxies at $z = 20$	21
2.8	Angular momentum module histogram for massive disc galaxies at $z = 20$	21
2.9	Over-density versus rotational maximum velocity.	22
2.10	Angular momentum versus rotational maximum velocity.	22
3.1	$R_s$ and $M_s$ versus $V_{\max}$ for massive disc galaxies in Ref-L0050N0752 simulation at $z = 20$	29
3.2	$M_s$ versus $R_s$ of massive disc galaxies in EAGLE's Ref-L0050N0752 simulation at $z = 20$	30

# List of Tables

1.1	Table with cosmological parameters and simulations properties used in this memory	3
1.2	Snapnum-Redshift-scale factor relation	5
2.1	Galaxy stellar mass function	14
3.1	Comparison of the number of galaxies in each simulation	24

# Chapter 1

## Introduction to EAGLE

### RESUMEN

*Este capítulo trata la introducción trabajo de Fin de Grado. Se compone por tres secciones: La sección 1.1, se describe brevemente el modelo físico en el que se encuentra enmarcado este estudio, así como, los aspectos más básicos a los que se hará alusión a lo largo de la memoria. El segundo, sección 1.2, se trata de una pequeña reflexión sobre la posible relevancia que tiene las simulaciones en el campo de investigación, en particular en la cosmología y astrofísica. Por último, la tercera sección (1.3) introduce al lector en el proyecto EAGLE, así como en sus aspectos más relevantes para esta memoria.*

The Evolution and Assembly of GaLaxies and their Environments (EAGLE<sup>1</sup> hereafter) is a Virgo Consortium's<sup>2</sup> project for cosmological supercomputer simulations, with a suite of hydrodynamical simulations which aims to understand how galaxies, and other cosmological bodies, form and evolve in the context of large-scale structure.

This computer calculations models are based in the standard cosmological model  $\Lambda$  *Cold Dark Matter* ( $\Lambda$ CDM for short) (Condon and Matthews [2018]) in a cube with a certain periodic size and mass with a certain finite resolution (i.e. particle number). This allows us to study the formation, assembly and evolution of cosmic structures and the dynamics of the intergalactic medium.

### 1.1 A brief visit to $\Lambda$ CDM

The current standard cosmological model is called  $\Lambda$  Cold Dark Matter ( $\Lambda$ CDM hereafter). It is a cosmological model currently accepted for being in accordance with observations (Planck Collaboration et al. [2014]), and fitting with the current conception of how the universe evolves. It is a particular model of the Friedmann-Lemaître-Robertson-Walker (FLRW hereinafter) universe paradigm. This type of universes are subject to the General Relativity Theory of gravity proposed by Albert Einstein (Einstein [1915]) as an exact equation for the Einstein's field equations.

This paradigm addresses the whole physical-theoretical framework needed to develop this thesis. Therefore, certain terms related to the model are inevitably alluded to throughout this paper and therefore worth a definition at the beginning of this thesis.

---

<sup>1</sup>See also <http://icc.dur.ac.uk/Eagle/>

<sup>2</sup>See also <http://www.virgo.dur.ac.uk/>

## Scale Factor ( $a$ )

In this model, as corroborated by observations (e.g. Perlmutter et al. [1999]), the universe is assumed to be expanding. This intrinsic property of the model is parameterised by the so-called scale factor. This is a dimensionless variable that expresses how much the universe has scaled (or descaled) with respect to a certain reference time ( $a \equiv a(t)$ ). By convention, the present time is used as a reference point, and the scale factor at the present time ( $t_{\text{present}}$ ) is taken as the unity,

$$a(t = t_{\text{present}}) \equiv a_0 = 1 \quad . \quad (1.1)$$

## Redshift ( $z$ )

Redshift is the displacement of the observed (obs) electromagnetic radiation with respect to the emitted (emit) radiation. It can be expressed as a function of frequency as the difference between emitted and observed frequencies divided by the observed frequency,

$$z = \frac{\nu_{\text{emit}} - \nu_{\text{obs}}}{\nu_{\text{obs}}} \quad . \quad (1.2)$$

In General Relativity, the theory on which the  $\Lambda$ CDM model is based, it is shown that an observer can measure a certain redshift when the emitter moves away from the observer. As the universe is in expansion, cosmological objects move away from each other, i.e. for these objects there is an intrinsic relation between the scale factor and the redshift,

$$a(z) = \frac{1}{1+z} \quad . \quad (1.3)$$

In short, in the type of universe that concerns us, expansionist, the redshift turns out to be a measure of both distance and type.

## Comoving Units

In a isotropic and expansionly system (due to the cosmological principle<sup>3</sup>), a set of privileged observer are the "comoving observers". These observers are the only ones that do not observe redshifting of light due to the expansion of the universe. Comoving coordinates are related to the proper coordinates ( $\mathbf{x}_p$ ) by the equation

$$\mathbf{x}_p(t) = \mathbf{x}_c(t) a(t) \quad . \quad (1.4)$$

Note that if for the proper observer a measured distance does not vary except for the expansion of the universe itself ( $\mathbf{x}_p(t) = \mathbf{x} a(t)$ ), for the privileged observer the measurement will not vary ( $\mathbf{x}_p(t) = \mathbf{x}$ ). Here is where the advantages of this type of coordinates lie.

Other interesting distance-related measures to be measured in a comoving way are velocity ( $\mathbf{v}_c$  and  $\mathbf{v}_p$  respectively)

$$\begin{aligned} \mathbf{v}_p(t) &= \frac{d}{dt}(\mathbf{x}_p(t)) = \frac{da(t)}{dt} \mathbf{x}_c(t) + a(t) \mathbf{v}_c(t) = \\ &= H(t) \mathbf{x}_p(t) + a(t) \mathbf{v}_c(t) \end{aligned} \quad (1.5)$$

---

<sup>3</sup>The cosmological principle may be described as a hypothesis which states that in the universe there is a homogeneous and isotropic distribution of matter at large (very large) scales.

Equations(1.8; 1.4) are taken into account in the last equivalence. Therefore, the last equality corresponds to two components, the first term makes reference to the velocity component due to the expansion of the universe while the second one is the so-called "peculiar velocity". Comoving and peculiar densities ( $\rho_c$  and  $\rho_p$  respectively) are related by the following equation:

$$\rho_p(t) = \rho_c(t) a^3(t) \quad (1.6)$$

Another observation, due to the convention taken with the scale factor value at present (1.1), the present time values of both position and density are the same for the proper observer and the comoving. More information of the comoving concept at The EAGLE team [2017] and Schaye et al. [2015] (Appendix D)

## Hubble Constant ( $H_0$ )

In the homogeneous and isotropic FLRW model, the velocity measured by Earth between any two comoving points (i.e.  $v_c = 0$ ) is (Mo et al. [2013], page 93)

$$\mathbf{v}_p(t) = \frac{da(t)}{dt} \mathbf{x}_c(t) = H(t) \mathbf{x}_p(t) \quad (1.7)$$

where  $H(t)$  is the time-dependent expansion of space-time term, the Hubble parameter.

$$H(t) \equiv \frac{1}{a(t)} \frac{d}{dt}(a(t)) \quad (1.8)$$

The Hubble constant is the present-day value of the Hubble parameter ( $H(t) = t_{\text{present}} \equiv H_0$ ), and this parameter is the employee in the Hubble-Lemaître law (Hubble [1929]). This empirical law is the observational relations between the distance ( $\mathbf{x}_p$ ) and the speed ( $\mathbf{v}_p$ ) of cosmological objects measured from Earth. This relation is linear and the proportionality factor is the Hubble constant ( $H_0$ ).

$$\mathbf{v}_p = H_0 \mathbf{x}_p \quad (1.9)$$

Name	Parameter	Value
Average density of matter	$\Omega_m$	0.307
Average density of dark energy	$\Omega_\Lambda$	0.693
Average density of baryonic matter	$\Omega_b$	0.04825
Hubble constant	$h \equiv H_0 \frac{s \text{Mpc}}{100 \text{km}}$	0.6777
Ref-L0050N0752 Comoving box size	$L$	50 cMpc
Ref-L0050N0752 number of DM particles <sup>4</sup>	$N$	$752^3$
Ref-L0050N0752 initial baryonic particle mass	$m_{\text{gas}}$	$1.81 \cdot 10^6 M_\odot$
Ref-L0050N0752 DM particle mass	$m_{\text{DM}}$	$9.70 \cdot 10^6 M_\odot$

Table 1.1: Table with cosmological parameters and simulations properties used in this memory



As discussed above, scale factor is related with redshift, so there is a relation between redshift and the Hubble constant, that it is as follows,

$$z = \frac{H_0}{c} x_p , \quad (1.10)$$

where  $c$  is the speed of light and  $x_p$  is the module of its vector ( $\mathbf{x}_p$ ).

Although the value of this parameter is currently under discussion (Di Valentino et al. [2021]), a value the one presented at the Plank Collaboration in 2013 (Planck Collaboration et al. [2014], Table 9)

The Hubble constant value is currently under discussion (Di Valentino et al. [2021]), in this paper it is taken as (Planck Collaboration et al. [2014]):

$$H_0 \simeq 67.77 \text{ km s}^{-1} \text{ Mpc}^{-1}$$

This is the value used in the EAGLE simulations that are analysed in this work (Schaye et al. [2015]).

### Critical Density ( $\rho_{crit}$ )

The critical density is one of the variables with more importance in this thesis. In the paradigm of the  $\Lambda$ CDM model. It is a guide value of the geometry of an FLRW Universe, its value is expressed (Mo et al. [2013], page 94) as

$$\rho_{crit} = \frac{3H_0^2}{8\pi G} \simeq 1.88 \cdot 10^{-26} h^2 \text{ kg m}^{-3} \quad (1.11)$$

where  $G$  is the gravitational constant and  $H_0$  the above-mentioned Hubble constant. In addition, another term relevant to this work is introduced, the density matter for the FLRW model, this parameter is usually defined as the ratio between the average matter density at present time ( $(\rho_p(t = t_{\text{present}}) \equiv \rho_0)$ ) and the critical density, therefore it is a constant parameter.

$$\Omega = \frac{\rho_0}{\rho_{crit}} \quad (1.12)$$

## 1.2 The Importance of Numerical Simulations.

Simulations are a research tool, a fundamental source of information for certain fields of research where it is difficult to have any way of validating or rejecting the veracity of the theoretical models.

The evolution of the large-scale distribution of matter is sensitive to a variety of fundamental parameters that characterise dark matter (henceforth DM), dark energy, and other aspects of our cosmological framework. The non-linearity of its process makes it impractical to describe them in an analytical way. For this reason, numerical simulations are crucial for understanding the behaviour of the models under research.

snapnum	z	a	snapnum	z	a	snapnum	z	a	snapnum	z	a
28	0.00	1.000	21	0.74	0.576	14	2.24	0.309	7	5.49	0.154
27	0.10	0.909	20	0.87	0.536	13	2.48	0.287	6	5.97	0.143
26	0.18	0.846	19	1.00	0.499	12	3.02	0.249	5	7.05	0.124
25	0.27	0.787	18	1.26	0.443	11	3.53	0.221	4	8.07	0.110
24	0.37	0.732	17	1.49	0.402	10	3.98	0.201	3	8.99	0.100
23	0.50	0.665	16	1.74	0.365	9	4.49	0.182	2	9.99	0.091
22	0.62	0.619	15	2.01	0.332	8	5.04	0.166	1	15.13	0.062
									0	20	0.047

Table 1.2: Equivalence between cosmological snapshot number, redshift and scale factor in EAGLE simulations (McAlpine et al. [2016], Table C.1)

In addition to the non-linearity of the processes, there are further reasons for simulations in the fields of cosmology and astrophysics.

Firstly, scalability, as cosmologists, it is impossible to study the predictions of the cosmological theories in a conventional laboratory because of the large scales of both time and space on which the models are found.

Secondly, observations. Because of the impossibility of studying in a laboratory, the only way to obtain data that can inform our models is through observation. On the one hand, from a particular point of view: the observation only shows a small amount of information of the total count, because only the current behaviour (for the observer) can be investigated. That is, if a galaxy is seen at a certain distance, it is impossible to observe the same galaxy at an earlier time. On the other hand, from a more general point of view, wanting to visualise the first moments of the universe means having to observe more distant bodies, with all the technological challenges that entails.

Numerical simulations make it possible to overcome the disadvantages mentioned above. The simulations can be replicated, which allows different physical models to be tested and those models that behave differently of the observations are discarded.

Although simulations have also problems with awaiting technological advances because there is a correlation between the computational time and the resolution of the simulation, several groups are making a great effort to advance both technologically and in the efficiency of mathematical models so that this relation (time-resolution) becomes smaller and smaller (e.g. Virgo Consortium).

This is where the EAGLE project appears as an state-of-the-art of hydrodynamic simulations in the field of galaxy formation. (Crain et al. [2015]; Schaye et al. [2015]). For this thesis, all the data necessary to carry out the tasks set out are extracted from this project. It is therefore essential to understand how this information is extracted from the simulations.

### 1.3 Inside EAGLE

The EAGLE simulations are based at the physical level on the standard cosmology model and at the computational level on a modified version of the public GADGET-2<sup>5</sup> code for simulations (Springel [2005]).

<sup>5</sup>See also <https://wwwmpa.mpa-garching.mpg.de/gadget/>

### 1.3.1 Initial Conditions and Cosmological Parameters

All simulations have certain constant parameters with a certain predefined value before starting the simulation. Some of these parameters are common to all simulations and others are specific to the simulation itself.

Others depend on each simulation. For this memory, the mass of each DM particle of the simulation (the rest of the particles have variable mass), the length of the simulated universe and the number of particles that are initially in the simulation are taken into account; Information from the last two is given in the suffix of the simulation name itself, which has the form LXXXXNXXXX, where instead of an X, it usually has a specific numbering. The first numbering of that suffix, the one after the letter L but before N, specifies the length (in comoving-Mpc or cMpc) of the simulation, while the second numbering specifies the cube root of the number of simulation particles per species, i.e. gas and DM (Schaye et al. [2015], Table 2).

Therefore, as an example, a simulation whose name has the following suffix L0050N0752 implies that it is a cube with a side size of 50 cMpc and  $752^3$  particles of both gas and DM type.

The name prefix also reports variations that are not taken into account in this work but allow calibration of the simulations available as detailed in Crain et al. [2015], page 7.

It should be noted that although these are the parameters mentioned in this memory, they are not the most important ones in the simulations, such as those parameters that define the numerical model of galaxy formation (the rate of star formation, the cooling of the gas, etc.). Such as the initial condition parameters of every simulation, for example: the square root of the linear variance of the matter distribution when smoothed with a top-hat filter of radius  $8 h^{-1}$  cMpc named  $\sigma_8$  ( $= 0.8288$ ), the scalar power-law index of the spectrum of primordial adiabatic perturbations ( $n_s = 0.9611$ ) or the primordial abundance of helium ( $Y = 0.248$ ) (Schaye et al. [2015], page 6).

The simulation analysed here is referred to as Ref-L050N0752, it is a simulation with  $2 \cdot 752^3$  particles in a cubic with a periodic volume of 50 cMpc. The initial mass of gas per particle is  $1.81 \cdot 10^6 M_\odot$  and the DM mass per particle is  $9.70 \cdot 10^6 M_\odot$  (Schaye et al. [2015], Table 2).

The set of important parameters used in this memory are listed in Table 1.1. The first four elements are cosmological parameters and equal for every simulation (Schaye et al. [2015], Table 1). the average densities are relevant to calculate the density of the region where a galaxy was formed at initial moments (explained in 2), while the Hubble constant is usually used as part of the units of the rest of properties in the *.hdf5* files (e.g. the coordinates of a particles a snapshot is stored with  $h^{-1}$  cMpc units; or its mass with  $h^{-1} M_\odot$ ). The following four correspond with parameters of the simulation used in this memory, it is referred to as Ref-L050N0752, it is a simulation with  $752^3$  DM particles, at snapshot 0 there are the same number of baryonic particles, in a cubic with a periodic volume of 50 cMpc. The initial mass of gas per particle is  $1.81 \cdot 10^6 M_\odot$  and the DM mass per particle is  $9.70 \cdot 10^6 M_\odot$  (Schaye et al. [2015], Table 2). This group of values are relevant to many aspects in the final part of this thesis, the calculation of the over-density and the angular momentum of the particles that are part of the galaxy but in the initial moments of the simulation (explained in section 2)

### 1.3.2 Snapshots

The EAGLE simulations run from a certain initial time ( $z = 127$ ), up to redshift zero. Over the course of the simulation, a total of 28 snapshots are saved of all simulation data at a certain time (i.e., at a certain redshift); the relation between the snapshot number (henceforth `snapnum`) and the redshift at which it is performed is specified in the table 1.2. All relevant information for these practices is contained in the files containing the data of the snapshots of the simulations.

### 1.3.3 Friends-of-Friends Algorithm

The Friends-of-Friends (FoF) method (Davis et al. [1985]) is an algorithm to identify groups of particles in a simulation such as they form an over-dense region. FoF is usually implemented in large-scale cosmology simulations to find structures and identifying features, in EAGLE case, it is used to find DM halos structures in the universe simulated (see Schaye et al. [2015])

#### GroupNumber and SubGroupNumber

A DM particle is linked directly to all other DM particles within a certain distance (its friends) and indirectly to all DM particles that are linked to its friends (its friends of friends). This creates networks of linked particles called groups. Each group reflects what in cosmology would be identified as an over-dense region of DM enclosed in a isodensity surface, a halo. Each group is assigned a number called `GroupNumber`, The smaller the `GroupNumber` implies that the halo is larger.

Then within the halo it is possible to determinate subhaloes running the SUBFIND algorithm (Springel et al. [2001]; Dolag et al. [2009]) using all particle species (gas, DM, stellar and black holes type). Galaxies can be determined and classified for their gravitational potential, for this classification a `SubGroupNumber` is assigned: the deepest gravitational well is the central galaxy (i.e. `SubGroupNumber = 0`) while the rest of galaxies, that are in the same halo, are its satellite galaxies (`SubGroupNumber > 0`).

Clarify that the `GroupNumber` and `SubGroupNumber` assigned to each formation for a certain snapshot does not necessarily hold for another snapshot, due to the dynamic behaviour of the universe.

### 1.3.4 How to Extract the Information

On the EAGLE webpage, in the *Public Data* section, the two ways in which the information can be extracted are specified, both of which are used throughout this research, so they will be explained briefly below. It should be noted that for both ways one must have a user account on the website.

#### EAGLE galaxy database

The first way allows us to obtain the galaxy information, determined by the FoF method, through a SQL format, to connect this information with a Python project using the *eagleSqlTools* library. All the information for extracting galaxy data is detailed in Stuart McAlpine’s article (McAlpine et al. [2016]).

## EAGLE particle data

The second way is to extract the simulation's own particle information by downloading the simulation's own *.hdf5* format files compressed into a *.tar* format. Although both formats are used in practice, the second one plays a greater role (see The EAGLE team [2017])

In this thesis, within the *.hdf5* files, we work with two sections. On the one hand, the *Header*, where all the constants and parameters (length of the universe, number of particles,  $h$ , scale factor, mass of the dark matter particles...) are collected. On the other side are the *PartType*, where all the characteristics of the particles themselves are collected (identification number, coordinates, velocity, typology, mass...), the *PartType* are four sections, each one of them has a suffix that specifies the type of particles (0 gas, 1 dark matter, 4 stellar, 5 black holes); so that if you wanted to look for information only about gas particles you would have to search specifically in *PartType0*.

# Chapter 2

## Objectives and Methodology

### RESUMEN

*En este segundo capítulo se nombran los objetivos de esta memoria. Además, se describe la metodología empleada para obtener los resultados necesarios para la discusión que se expondrán en el siguiente capítulo. En la metodología se explica los pasos hasta obtener la función de distribución de la velocidad de rotación de galaxias masivas de tipo disco de la simulación Ref-L0050N0752 (sección 2.2), y como se describe cómo se obtienen los datos de sobredensidad y momento angular para comparar con la función de distribución (sección 2.3).*

### Objectives

The aim of this thesis is to investigate the maximum rotational velocity distribution function of massive disc-type galaxies from a particular EAGLE simulation (Ref-L0050N0752) at  $z = 0$  (snapshot 28) and study whether any relation between this distribution and the environment where galaxies can be formed.

To this end, on the one hand, the field of cosmological physics is introduced from a quantitative point of view, a vision connected to both the observational and the theoretical aspects of this field. On the other hand, with a little more distance from the academic aspect, through the objectives of this thesis, one can extract lessons and tools necessary for research and even for personal organisation.

### Methodology

The methods employed in this work are described here. First, it is explained how to select the disc galaxies in the EAGLE database; with the galaxies sample determined, the maximum rotational velocity of galaxies and their stellar mass are used to derive their distribution, with a previous test to verify their Tully Fisher Relation. Then the code used to obtain the over-density and angular momentum introduced in the objective is described, and finally the obtained data is presented. All the findings of this chapter will then be discussed in the next chapter, followed by a conclusion.

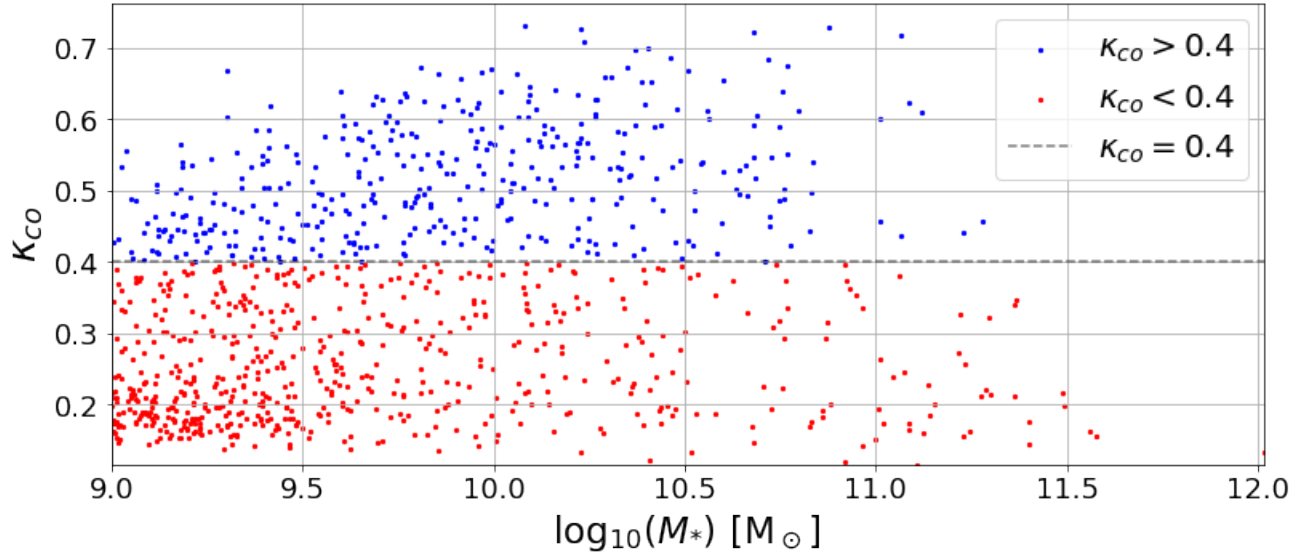


Figure 2.1:  $k_{co}$  versus  $M_*$  for massive galaxies ( $M_* > 10^9 M_\odot$ ) in Ref-L0050N0752 at  $z = 0$  (snapshot 28). With the data obtained and thanks to the conditions imposed, it can be affirmed that this simulation contains 990 massive galaxies (385 of them disc) at  $z = 0$ .

## 2.1 Morphology of Disc Galaxies

The galaxies of interest in this work are the spiral galaxies, which are usually flat and most stars are formed in a rotating disc. For this reason, they are usually referred to as disc galaxies.

Disc galaxies have many mutually dependencies, which are related to each other. Regarding morphology (e.g. Graham et al. [2018]), with the colour (e.g. Blanton et al. [2003]; Driver et al. [2006]), luminosity and rotation velocity (e.g. Tully and Fisher [1977]), internal kinematics, morphology and mass (e.g. Dressler [1980]; Baldry et al. [2006]; Kelvin et al. [2014]), star formation rate (e.g. Kennicutt [1983]) and more.

The classification of galaxies in EAGLE simulations by their morphology is a topic of current study (e.g Avila-Reese et al. [2008]; Correa et al. [2017]; Thob et al. [2019]). Correa et al. make a detailed study of how the morphology of the galaxies simulated in EAGLE relates to the colour and with the cosmological observations made. They find that it is possible to differentiate between the two basic morphology types of galaxies by the parameter defined as  $\kappa$  *corotating* (or  $\kappa_{co}$ )

This parameter specifies the fraction of a particle's total kinetic energy that is invested in ordered co-rotation (more details Correa et al. [2017]). If  $\kappa_{co} > 0.4$  the galaxy can be defined as a disc-type galaxy, otherwise it is of elliptical type. This parameter is included in the characteristics of the galaxy provided by EAGLE. It should be noted, however, that galaxies are usually neither one type nor the other, but a combination of both.

In this thesis, in addition to being disc galaxies ( $\kappa_{co} > 0.4$ ) there are two more conditions, they have to be massive and the central galaxy in their group. Therefore, it will also be taken as a condition

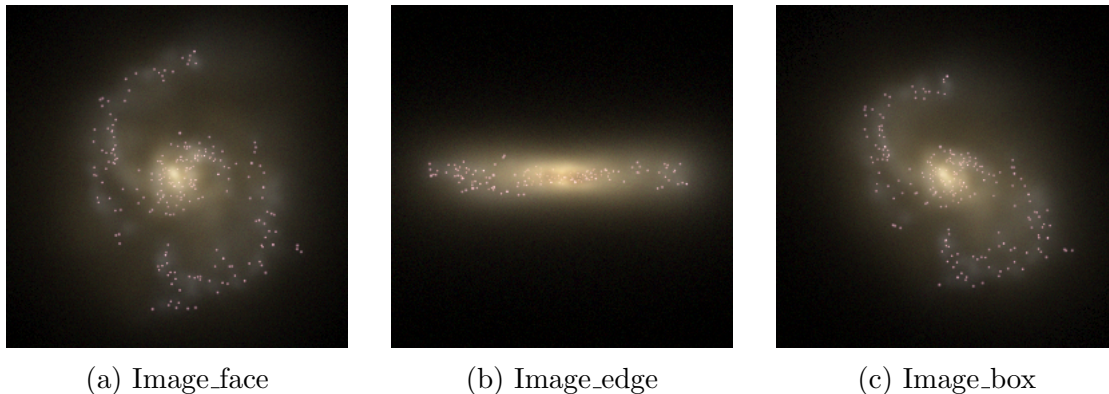


Figure 2.2: Mock gri images of a galaxy at  $z = 0$  as available in the database (McAlpine et al. [2016]). The images are 60 pkpc on a side. Each image represent a different perspective (the perspective names are based in the database nomenclature) of the same simulated massive galaxy (GalaxyID = 1479010 in the Ref-L0050N0752 simulation), the galaxy is who has the higher  $\kappa_{co} (\simeq 0.73)$ , so it can be considered a disc-type galaxy. It can be seen that the galaxy is indeed a disc galaxy when viewed "face-on", while when viewed "edge-on" it is flat.

that the galaxies to be studied must have a stellar mass ( $M_*$ ) higher than  $10^9 M_\odot$  (Figure 2.1) and its SubGroupNumber must be zero. The choice of these conditions allows us to have a good set of galaxies to study. The choice of the main halo galaxy is due to the fact that we try to study galaxies whose behaviour is not compromised mainly by dynamical interaction with galaxies that have a larger gravitational well, trying to obtain data with as little distortion as possible.

## 2.2 Rotational Velocity Distribution Function

Once the candidate galaxies are selected for study, it is time to determine the distribution function for the rotation of this set of galaxies. This distribution function is a probability distribution subject to a certain law not determined in this work. For this purpose, the rotation speed of the galaxy is taken as the parameter to obtain this information.

It is well known that the rotation curve of disc galaxies is not constant over the whole radius (e.g. Freeman [1970]; Rubin and Ford [1970]; Sofue and Rubin [2001]), indeed, the way in which the rotation of the galaxy varies is used as an argument for the existence of DM (a current review of this field: Freese [2009]). This implies that one has to be precise in choosing the parameter defining the rotational velocity of a galaxy.

The property of the simulated galaxies  $V_{\max}$  (defined in Crain et al. [2015] and used in Schaye et al. [2015]) is taken as the reference for this work, it is provided by the EAGLE project and defined as the maximum of the circular velocity rotation curve.

$$V_{\max} = \max \left( \sqrt{GM(< r)/r} \right) \quad (2.1)$$

where  $G$  is the gravitational constant,  $r$  is the selected radius and  $M(< r)$  is the mass within the radius. As this choice is arbitrary, and drawn from theoretical approaches, before studying the



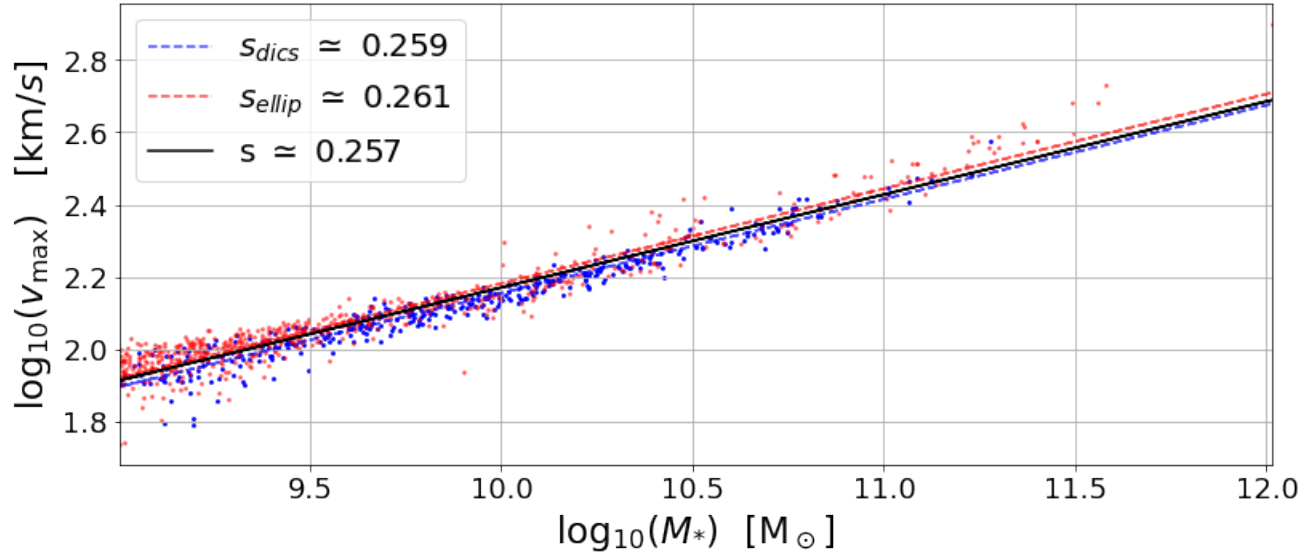


Figure 2.3:  $V_{\max}$  versus  $M_*$  for massive galaxies ( $M_* > 10^9 M_\odot$ ) in Ref-L0050N0752 at  $z = 0$  (snapshot 28). Linear point fits have been calculated for all galaxies (black), for disc-type galaxies (blue) and for ellipsoid-type galaxies (red). The slope of each of the fits is shown in the legend of the figure.

distribution function of this parameter, we make sure that it at least correlates with the other properties (e.g. stellar mass).

### 2.2.1 Tully-Fisher Relation

As mentioned above, disc galaxies have global scaling relations. The well-known Tully-Fisher relation (TFR; Tully and Fisher [1977]) is one of them. TFR is an empirical scaling relation between the rotation velocity and luminosity of a disc Galaxy.

Because of the correlations of the properties of a disc galaxy (between colour, rotation speed, luminosity, baryonic mass, stellar mass...), the TFR can be extrapolated to other correlations (e.g. Pierce [2000]; Stone et al. [2021]).

In this work the maximum rotational velocity  $V_{\max}$  is compared versus the stellar mass ( $M_*$ ), and it is verified whether they show a power law, which can be referred to as stellar TFR (or S-TFR), ( $M_*$ ) (see Avila-Reese and Firmani [2001]), given as:

$$\log_{10}(V_{\max}) = s \log_{10}(M_*) + c \quad (2.2)$$

where  $s$  is the slope and  $c$  is the y-intercept. This fit is shown in Figure 2.3. The result (discussed in the next chapter) of this relation may help to determine whether  $V_{\max}$  can be taken as a suitable parameter for the rotation speed of galaxies.

TFRs are topical studies (e.g. Ferrero et al. [2017]; Stone et al. [2021]). It is a good field from which to extract a lot of information about disc galaxies: the evolution of galaxies, their formation and their dynamics. Having a larger catalogue of galaxies with more precise data on their properties will allow us to verify the good behaviour of the model studied.

## 2.2.2 Distribution Function

Given a data set of a certain parameter, the probability of obtaining a certain expected value (or range of values) in that environment can be characterised by a probability distribution associated with that variable.

With the sample of galaxies chosen and determined  $V_{\max}$  as the rotational velocity term,  $V_{\max}$  and  $M_*$  distributions are shown in the Figure 2.4, apart from some comparisons. For this purpose, four graphs are plotted with histograms of the variables of interest.

Note the image on the top right, the histogram is made about the  $V_{\max}$  of the disc massive galaxies of Ref-L0050N0752 simulation, it describes the probability that a massive disc galaxy has a certain rotational velocity in the Ref-L0050N752 simulation. However, this and the other images will be discussed in 3.5

Apart from histograms, there are other standardised formalisms to derive the mass distribution function of galaxies (see Weigel et al. [2016]). One popularised method is the density-corrected maximum volume (Schmidt [1968]), which allows one to express the value of each of the objects (in this case the stellar mass of the galaxies) weighed by the maximum volume at which they could be detected. The results are usually expressed as the number density of galaxies, in this case  $\frac{dn}{d\log_{10}(M_*)} [\text{cMpc}^{-3}]$  (where n is the galaxy number). This allows studies involving measurements from different samples (e.g., Li and White [2009]; Baldry et al. [2012]; Bonfini et al. [2021]) and from simulations (e.g., Schaye et al. [2015]; McAlpine et al. [2016]) to be compared. In addition, expressing on the number density of galaxies allows one to study and compare with theoretical well described fits (e.g. the Schechter functions Schechter [1976])

The EAGLE papers studied for this thesis show the functions as a function of galaxy number density (e.g. Crain et al. [2015], Figure 2). The work of McAlpine et al. [2016] details, in Appendix A, a code with which to replicate the GSMFs (dex = 0.2) obtained by Schaye et al. [2015] in their Figure 4. This same code is used in this study to obtain the GSMF of the massive disc galaxy sample of the EAGLE simulation Ref-L0050N0752 at  $z = 0$  (Data points in Table 2.1, and plotted in black curve of the bottom-right image in Figure 2.4 and orange curve in Figure 2.5), it is also used for the galaxy stellar mass function of all massive stars of the same simulation and same redshift (black curve in Figure 2.5) and replicating the curve of Ref-L100N1504 obtained by Schaye et al. [2015] in the already mentioned Figure 4 (Figure 2.4)

## 2.3 Properties under Initial Conditions.

As described in the objectives section, after the distribution function of  $V_{\max}$  at  $z = 0$  it is necessary to study whether it is related to certain parameters of the initial moments of the simulation,  $z = 20$ . The parameters chosen for the initial conditions are the over-density and the angular momentum of the region in which the particles of each of the galaxies are found at snapshot 0.

$\log_{10}(M_*) [M_\odot]$ mid point	$dn/d\log_{10}(M_*)$ [ $10^{-3} dex^{-1} cMpc^3$ ]	bin width	number
9.15	62	0.2	35
9.35	72	0.2	41
9.55	79	0.2	38
9.75	85	0.2	63
9.95	98	0.2	52
10.2	81	0.2	44
10.4	50	0.2	50
10.6	35	0.2	25
10.8	8	0.2	25
11.0	5	0.2	4
11.2	2	0.2	6
11.4	1	0.2	2

Table 2.1: Galaxy stellar mass function for disc massive galaxies of Ref-L0050N0752 simulations

To do so, we modify a code written by the author of this thesis in *Prácticas Externas* (an external internship university subject), which was described in the memory and defended in front of the tribunal afterwards. This code is written in the Python programming language and was designed to obtain the over-density of a region at  $z = 20$  delimited by the particles of a galaxy selected at  $z = 0$  for the simulation Ref-L0012N0188.

### 2.3.1 Internship Code Modifications

The code developed in the internship is an algorithm, which behaves as follows:

1. Loading snapshots 0 and 28.
2. Selecting a galaxy at snapshot 28 ( $z = 0$ ).
3. Saving the ParticleIDs.
4. Looking for this particle set at snapshot 0 ( $z = 20$ ).
5. Looking for its center of mass.
6. Looking for the further<sup>1</sup> particle from the center of mass.
7. Calculating its distance.
8. Calculating the volume and the complete mass inside the volume.
9. Calculating the density and over-density.

The simulation Ref-L0012N0188 is the smallest simulation of the whole EAGLE project that is usually used for testing. For the simulation used in this thesis, Ref-L0050N0752, the code is hardly modified.

Firstly, it is decided to work remotely with the computer of the *Galaxy Evolution Theory research group*<sup>2</sup>. Due to the large amount of data stored in each of the snapshots which can be found at around 30 and 60 GB each one, it is decided to run the program on the group's computer, a

<sup>1</sup>The one furthest from the 90% closest to the centre of mass.

<sup>2</sup>See also <https://www.iac.es/en/projects/numerical-astrophysics-galaxy-formation-and-evolution>

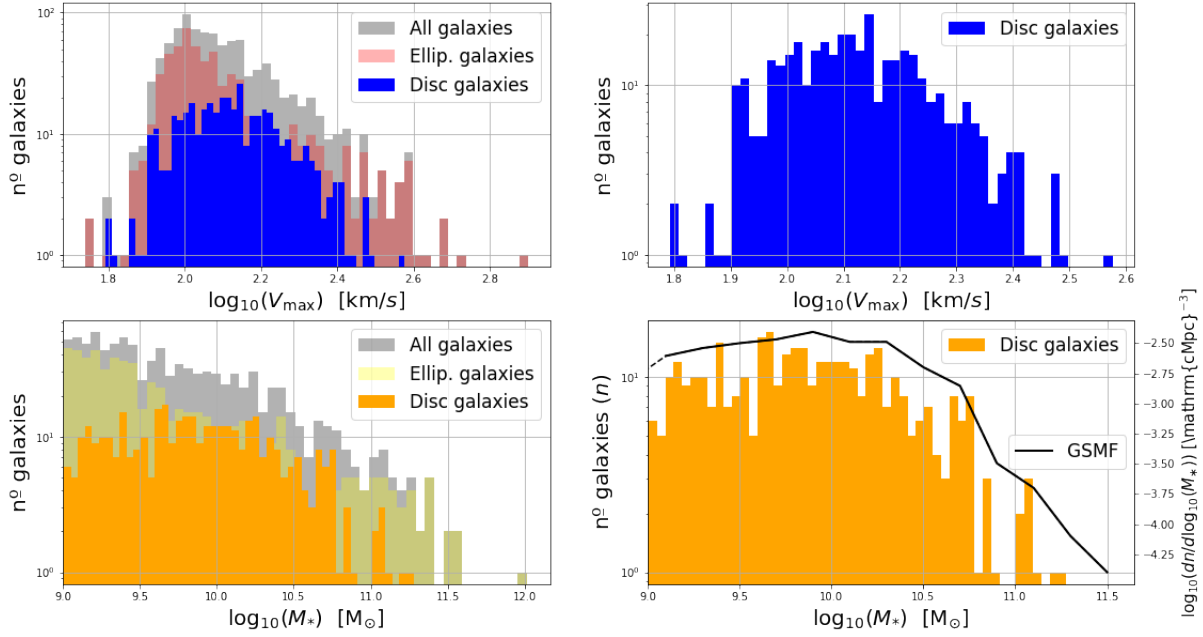


Figure 2.4: The first image (top-left) shows the distribution for all the galaxies shown in Figure 1, the total set is black, elliptical galaxies are red, and disc galaxies are blue. The second image (top-right) focuses only on the disc group. The third image (bottom-left) is analogue to the first but for  $M_*$ , such that the orange is for disc galaxies, the yellow represents the elliptical set and newly the black is for the total set. The last image (bottom-right) has two representations, the histogram, that is like the second one focused on the disc group of massive galaxies, and the galaxy stellar mass function (GSMF) .

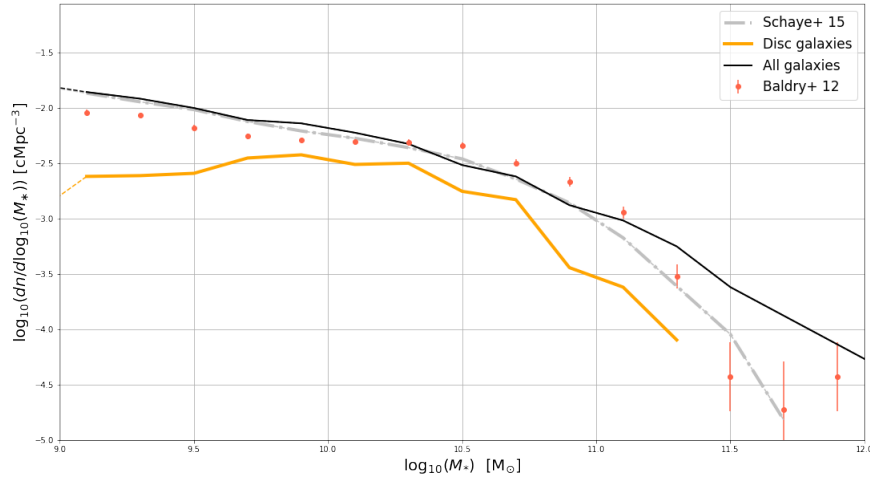


Figure 2.5: Comparison of galaxy stellar mass functions. The first curve (silver) replicates the Ref-L100N1504 GSMF (at  $z = 0.1$ ) of Schaye et al. [2015], based in McAlpine et al. [2016] Appendix A code. The second curve (orange) is shown in Figure 2.4 in the bottom-right image and corresponds to the GSMF for disc massive galaxies. The third curve (black) is the analogous galaxy stellar mass function but for the whole massive galaxies sample, refers to the grey histogram of the top-left image of Figure 2.4. Finally, data points (red) are part of the Baldry et al. measurements (Baldry et al. [2012], Table 1) at  $z < 0.06$ .

computer with sufficient specifications to run the corresponding snapshots of the simulation (first step of the algorithm).

Secondly, the code first is modified and then rewritten to be able to work remotely and to be efficient enough for the computer to run the program. (third to eighth). This is due to group's computer did not be able to run completely the code and stopped it. Thanks to Dr. Dalla's help and for his advice, it is possible to modify the most critical parts so that, at least, the computer is able at least to run the whole program.

One of the main changes in this aspect is to replace functions that need to duplicate arrays in order to modify it (e.g. `append` or `concatenate` functions from the *NumPy* library<sup>3</sup>). To this end, other functions are written to obtain the same result but without duplicating the selected information, these functions are achieved because one already knows the number of elements to be used. Then, several modifications are made with the aim of making the calculation faster, improving the performance per galaxy by 50%, which is equivalent to saving days of waiting to obtain the necessary information.

It should be noted that there are still some bottleneck steps (especially in the fourth step of the algorithm with the `in1d` function of the *NumPy* library), which could be studied to improve the code and thus allow studies with larger simulations to be carried out in a reasonable calculation time.

Finally, the code is modified again so that it not only provides information on the over-density (the ninth algorithm step) but also information on the angular momentum, which is of interest for this thesis. Therefore, the relevant data to calculate these parameters is saved in a *.hdf5* file, the same format as the data of each snapshot.

### 2.3.2 TFG Code Algorithm.

The new code presents more than 500 lines of code condensed into one class and 3 functions (Appendix A):

- The *Data\_snapnum* class. This class allows the user to extract the data from the *.hdf5* files of a certain Snapshot efficiently and with the correct format of the output to use in functions afterwards. The class is mainly based on the *read\_dataset* function described in the work developed by The EAGLE team [2017] (section 4.1) and the rest are extra functionality around *read\_dataset* to match output to function inputs.
- The function *Galaxy\_to\_past*. This is a function that allows us to determine the region at the beginning of the simulation that must be taken into account in order to study the properties of this region selected. The output of this function corresponds to obtaining the centre of mass and the radius of the sphere, which is equivalent to steps 5 and 6 of the algorithm described in the previous section (2.3.1).

---

<sup>3</sup>See also <https://numpy.org/>

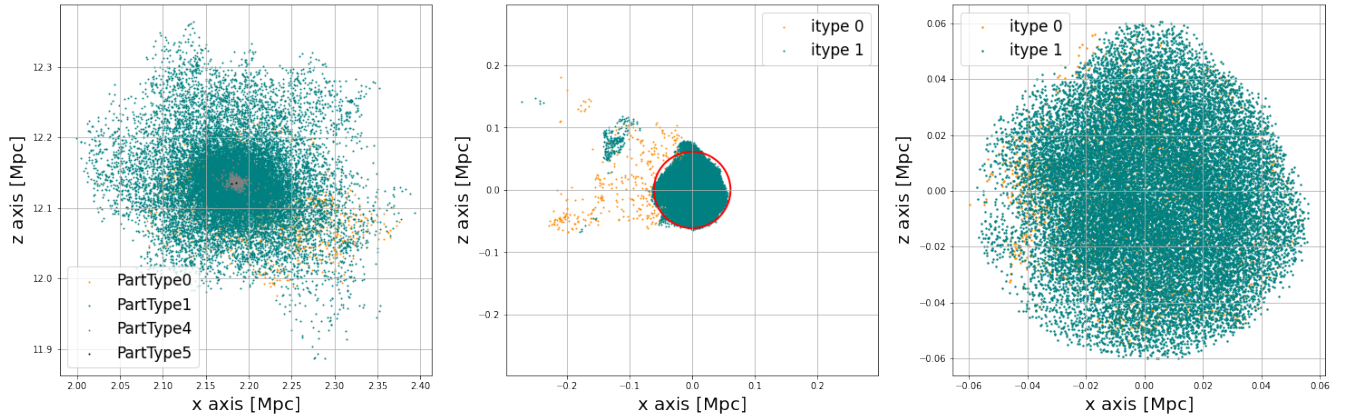


Figure 2.6: Representations of some of the phases of the modified code of the practice period. The image on the right is a two-dimensional representation of the galaxy with ID 59467 found in snapshot 0 of the simulation Ref-L0012N0188 of the EAGLE project, showing the position of each of the particles, as well as the type. The image in the centre shows the same particles as in the first image but at snapshot 0, with a red circle informing the perimeter where 90% of the particles are inside. The image on the right is the same as the previous one but only with the particles that were inside the perimeter.

- The function *Overrho*. This is the function to obtain the over-density of the region determined by the function *Galaxy\_to\_past*, equivalent to step 9 of section 2.3. This section will discuss this function in more detail.
- The function *AngularMomentum*. This is the function that allows us to obtain the angular momentum of the region determined by the function *Galaxy\_to\_past*. This is a completely new function developed from scratch. It will be discussed in more detail.

With all the modifications implemented, the code (Appendix A) used to obtain the angular momentum and over-density characteristics at the initial moments of the simulation is as follows

1. Loading snapshots 0 and 28. From the public database provided by the EAGLE project (login and password required) it is possible to obtain the data packets of each snapshot of any simulation (see for details The EAGLE team [2017]), these packages will be needed to load the particle data from the snapshot of interest.
2. Selecting a galaxy at snapshot 28 ( $z = 0$ ). Using the class called *Data\_snapnum* is possible to show the galaxy catalogue of the determined snapshot, that catalogue can be obtained directly from the website in SQL format (see for details and McAlpine et al. [2016]). From the catalogue the IDs of the galaxies (GalaxyID) necessary is extracted this thesis, i.e. massive ( $M_* > 10^9 M_\odot$ ), the principle galaxy of the whole halo (SubGroupNumber = 0) and disc ( $\kappa_{co} > 0.4$ ) galaxies.

In addition to the GalaxyID, other parameters can also be extracted from the galaxies catalogue, such as their mass (stellar or total), their  $V_{max}$ , their  $K_{co}$ , their illumination, their colour, etc. In fact, it is in this way that the information necessary to create the Figures 2.1, 2.4 and 2.3 is obtained.

3. Saving the DM ParticleIDs. From the GalaxyID the GroupNumber and SubGroupNumber of each galaxy can be obtained. Then a mask of the particles found at snapshot 28 in the galaxy can be created, and a list with the ID of each particle (ParticleIDs) can be obtained (Figure 2.6, first image). Unlike the GalaxyIDs, the ParticleIDs are the same during the whole simulation.

One of the modifications made to improve the performance of this code is to search only for DM particles (ParticleType1). With this measure, the number of particles to be looked for in snapshot 0 is considerably reduced. In addition to the fact that the centre of mass obtained does not change much and it is quite likely that the particles farthest away from the centre of mass are of type DM, since the rest of the particles are understood to be closer to the halo core.

4. Looking for this particle set at snapshot 0 ( $z = 20$ ). Once the ParticleIDs in snapshot 28 are known, the *Galaxy.to.past* function is used to create a mask that can be generated indicating where they are located within the entire list of particles in snapshot 0. As mentioned in the previous section, this may be the worst implemented part of the code, it is the section that takes the longest to calculate and where major modifications could be made to improve the performance of the program.
5. Looking for its center of mass ( $r_{cm,j}$ ). With the created mask the positions of all selected particles can be obtained. As these are DM-type particles, the mass of these particles is equal and constant for all of them during the whole simulation (Schaye et al. [2015], Table 2), so the centre of mass ( $M_s$ ) can be calculated.

$$\mathbf{r}_{cm,j} = \frac{1}{M_j} \sum_k m_k \mathbf{r}_k \quad (2.3)$$

where  $m$  is the particle mass and  $\mathbf{r}$  their coordinate vector. The  $j$  index means the  $j$ -st galaxy of the sample while the  $k$  index is about the particle set with the ParticleID determined in the third step.

6. Looking for the 90% further particle from the center of mass. Knowing the centre of mass and the positions of the particles, a radius ( $R_s$ ,  $s$  comes from the sphere) can be calculated and thus the volume covering all particles can be calculated. In the internship, it was observed that doing this could cause problems due to particles that behave abnormally and are strangely far from the halo, which would cause the volume to be over-weighted and its density to be underestimated.

During the internship period, the causes of these anomalous behaviours were not explored in depth, but some reasons were proposed. Within this set of ideas, the most widely accepted hypothesis is to blame these anomalies on the low quality and small size of the simulation Ref-L0012N0188. This would mean that the distribution is not ideal and that there could be particles that are far enough away that they are not part of the same core, but in the future they will be, because these "satellite" particles have no gravitational well to collapse into. In any case, this is open to study.

As a possible solution, we took the 90% of the particles that are closer to the centre of mass, with the idea of skipping these possible anomalous particles. This solution may underweight the volume, but not enough to make the results unreliable. In any case, this could be another aspect to be modified in case we want to make progress in this research.

As an example, in the second image of 2.6 is possible to see how there is a kernel of particles but there are some particles that would make the sphere inappropriately larger, in contrast, the approach at 90% may be a good measure of the volume of the kernel (third image).

7. Calculating the volume ( $V_s$ ) and the complete mass ( $M_s$ ) inside the volume. With the variables calculated in the previous step as input of *Overrho*, the volume occupied by a sphere with a radius equal to the distance obtained can be obtained. Moreover, since the position of the centre of mass and that of all the particles in the snapshot are known, it is possible to make sure which particles are inside the sphere and calculate the total mass inside it.

$$V_{s,j} = \frac{4}{3} \pi R_{s,j}^3 \quad (2.4)$$

$$M_{s,j} = \sum_{\substack{i \\ \forall r_i < R_{s,j}}} m_i \quad (2.5)$$

where  $r$  is the module of the coordinates of a particle and the  $i$  index is used for the  $i$ -st particle of the whole set of particles in the universe.

8. Calculating the density ( $\rho$ ) and over-density ( $\Delta\rho$ ). With mass and volume, the density of the determined region can be obtained trivially.

$$\rho_j = \frac{M_{s,j}}{V_{s,j}} \quad (2.6)$$

In addition, the average material density ( $\rho_0$ ) of the universe can be calculated. First obtaining the critical density, which can be calculated from the hubble constant (a constant parameter within the simulation, see Table 1.1) 1.11. Then, the material density is another constant within the simulation, (see Table 1.1), thus it is possible to use 1.12 and substitute the average density. It shall read as follows

$$\rho_0 \simeq 1.88 \cdot 10^{-26} \Omega h^2 \text{kg m}^{-3} \quad (2.7)$$

The over-density is given as

$$\Delta\rho_j = \frac{\rho_j - \rho_0}{\rho_0} \quad (2.8)$$

9. Saving over-density information. With the help of the Python *h5py* library<sup>4</sup>, an *.hdf5* file with a data table is created with the following data: the position of the centre of mass, the distance to the furthest particle (from the 90% of the nearest particles), the density of that region, the over-density of the region.
10. Calculating and saving the angular momentum ( $\mathbf{L}$ ). The other, and new for the code, property of the original formation region of the galaxies sample to compare with  $V_{\text{max}}$  is the total angular momentum. This physical magnitude is a vector measurement that characterises the inertial properties of a body, which is rotating relative to a certain point, the angular momentum is given as

$$\mathbf{L}_j = \sum_{\substack{i \\ \forall r_i < R_{s,j}}} m_i \mathbf{r}_i \times \mathbf{v}_i \quad (2.9)$$

where  $m$  is the mass,  $\mathbf{r}$  the coordinate vector and  $\mathbf{v}$  the velocity vector. The  $i$  index indicates the  $i$ -st particle of the set of particles that are inside in the volume defined in the seventh step of this algorithm and the  $j$  index represents the  $j$ -st galaxy of the sample. The vector product in the code is developed as follows:

$$\mathbf{r}_i \times \mathbf{v}_i = [(r_{1,i} v_{2,i} - r_{2,i} v_{1,i}), (r_{2,i} v_{0,i} - r_{0,i} v_{2,i}), (r_{0,i} v_{1,i} - r_{1,i} v_{0,i})] \quad (2.10)$$

---

<sup>4</sup>see also <https://www.h5py.org/>



Where the first subindex indicates the position within the vector. Since we are dealing with three dimensions and we are working with the Python programming language, the subindex can be (0, 1, 2). As the code had already obtained both the centre of mass and the radius of the volume of the region to be studied for each galaxy (in the seventh step of this algorithm), in this part it only remains to identify the particles that are in that region, extract their properties necessary to perform this calculation (i.e. their coordinate vector, their velocity and their mass) and calculate the angular momentum vector. This vector is also stored within the same dataset, in the file *.hdf5*, with the rest of the data, in order to fulfil the third part of this work.

As the purpose of obtaining these values is to compare with  $V_{\max}$ , the modulus of the angular momentum vector is used for this purpose. The modulus is calculated as follows:

$$L_j = \left( (L_{0,j})^2 + (L_{1,j})^2 + (L_{2,j})^2 \right)^{\frac{1}{2}} \quad (2.11)$$

### 2.3.3 Over-density and Angular Momentum

With the code explained in the previous section the over-density and angular momentum data are obtained for the region where the galaxies in our sample formed at  $z = 20$  (snapshot 0). To conclude this chapter, data obtained for these properties are shown, as a histogram, in Figures 2.7 and 2.8 respectively, and their comparisons with the  $V_{\max}$  of the same sample of galaxies in Figures 2.9 and 2.10. The following chapter discusses the results obtained.

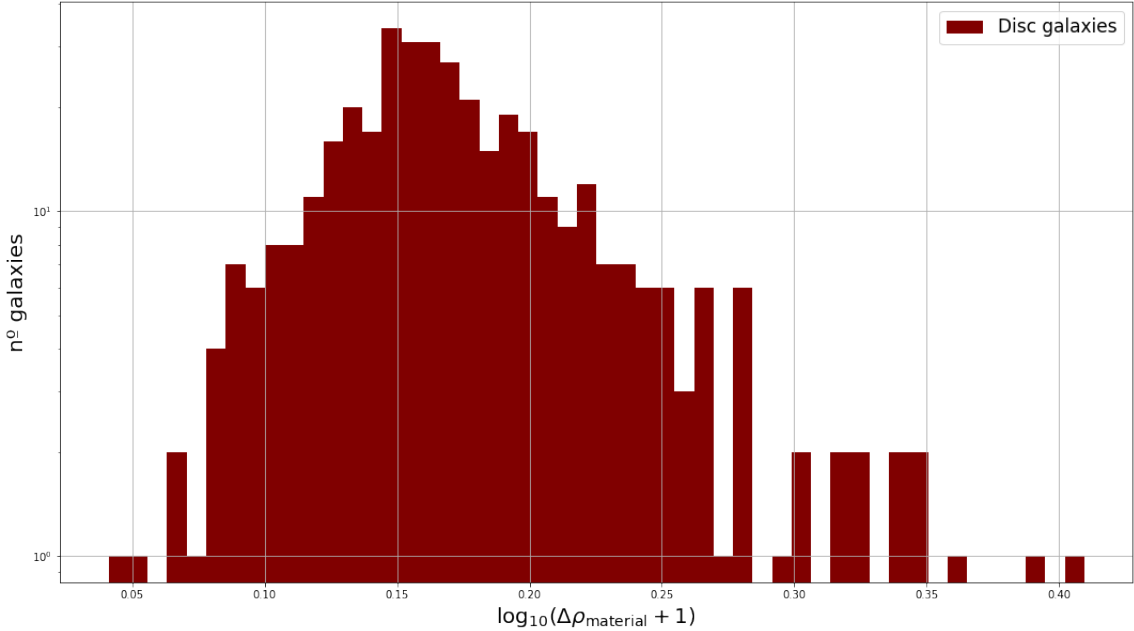


Figure 2.7: Over-density histogram for massive disc galaxies of the Ref-L0050N0752 EAGLE simulation at  $z = 20$ .

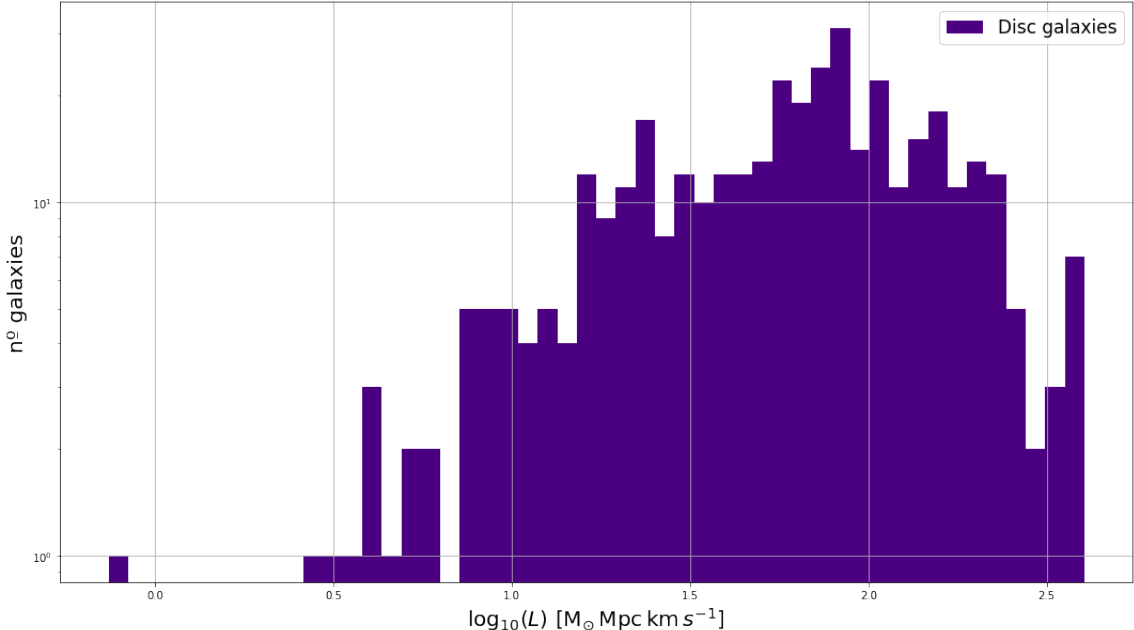


Figure 2.8: Angular momentum module histogram for massive disc galaxies of the Ref-L0050N0752 EAGLE simulation at  $z = 20$ .

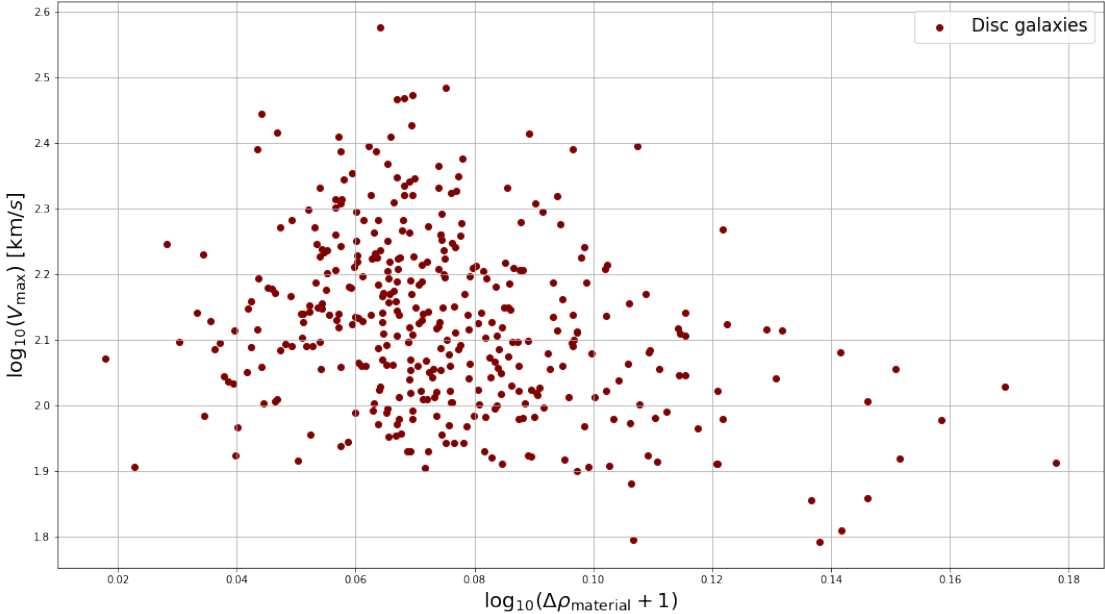


Figure 2.9: Over-density versus rotational maximum velocity. Comparison between the over-density in the region where galaxies form at  $z = 20$  and the  $V_{\max}$  property of massive disc galaxies at  $z = 0$  (EAGLE’s Ref-L0050N0752 simulation).

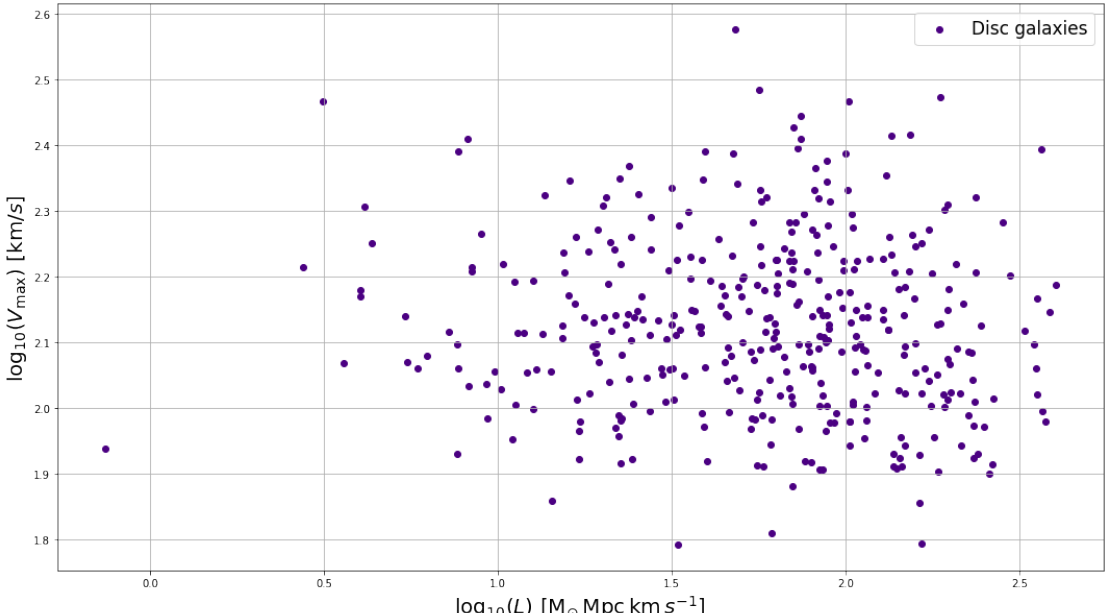


Figure 2.10: Angular momentum versus rotational maximum velocity. Comparison between the total angular momentum in the region where galaxies form at  $z = 20$  and the  $V_{\max}$  property of massive disc galaxies at  $z = 0$  (EAGLE’s Ref-L0050N0752 simulation).

# Chapter 3

## Discussion of the Results and Conclusion

### RESUMEN

*El tercer capítulo cierra la memoria de este trabajo de fin de grado. Se trata de la discusión de los resultados obtenidos a lo largo del estudio siguiendo el orden establecido en la metodología (2). Se comienza (sección 3.1) valorando si la simulación seleccionada resulta ser lo suficientemente buena como para usarla para el estudio probabilístico. Luego se valora la propiedad  $V_{\max}$  de las galaxias en la sección 3.2 a través de los resultados obtenidos en la TFR. En la sección 3.3 Se describe las funciones de distribución obtenidas (sección 2.2), así como la comparación entre ellas. La última discusión de resultados (sección 3.4) se divide en dos partes, primero sobre el código reescrito en este trabajo y luego con los resultados obtenidos por dicho código. Se finaliza el capítulo con una conclusión de todo lo mostrado en este trabajo.*

### 3.1 Ref-L0050N0752 Simulation

The EAGLE simulation Ref-L0050N752 has been used for this final degree project to carry out statistical studies. The Ref-L0050N0752 presents a high resolution simulation, due to the number of particles. It also presents an intermediate size (50 cMpc), there are 100, 50, 25 and one case of 12.5 cMpc, (for further information, see Crain et al. [2015] ).

There are several reasons for the choice of this simulation, all due to the fact that for this work it is not possible to choose superior simulations, e.g. Ref-L0100N1504 (see other possible simulations in Crain et al. [2015]). The reasons include the number of galaxies that can be used as samples and the computational demands required to obtain all the necessary data. it presents a sufficiently large sample of galaxies to be able to perform a decent study of the aspects addressed in this work, i.e. 385 disc galaxies (Table 3.1). Comparing the number of galaxies used in this study with the rest of the simulations shown in the table, it can be seen that, although not ideal, it is an appropriate set to investigate. Although it is left as a precedent for a possible study for a higher simulation. This simulation turns out to be a good balance large enough for the statistical data to be relevant but large enough for the quality of the data to be adequate. This synergy allows to have good results for this research.

It must be understood that with the increase in both properties (the size and resolution of the universe), it is not just that there are more galaxies that match the ones we are looking for, but they also have a higher resolution, i.e. a higher number of particles per galaxy. Therefore, the scale of computational demand when switching from one simulation to other higher is not linear to the number of galaxies, but also scalates with the resolution of each galaxy. Added to this are the bottlenecks in the code, such as possible inefficiencies that have not yet been detected.

Simulation	Number of galaxies	Number of massive galaxies	Number massive disc galaxies
Ref-L0012N0188	47	21	5
RecalL0025N0376	349	135	52
Ref-L0025N0752	900	166	78
Ref-L0050N0752	2838	900	385
Ref-L0100N1504	21445	7482	2830

Table 3.1: Comparison of the number of galaxies in each simulation. This shows the number of total galaxies. Then which of them are massive ( $M > 10^9 M_\odot$  and `SubGroupNumber = 0`). And finally, within the second group which are of disc type ( $\kappa_{co} > 0.4$ ).

## 3.2 Tully-Fisher Relation

The stellar TFR is predicted, with the rest of the scaling relations of disc galaxies in the current theory of galaxy formation and evolution based on the  $\Lambda$ CDM theory (see e.g. Mo et al. [1998]; Gnedin et al. [2007]; Courteau et al. [2007]). Not only that, it has also been studied observationally (e.g. Stone et al. [2021]), and compared with the theoretical models (e.g. Bell et al. [2003]; Trujillo-Gomez et al. [2011]), obtaining positive results. Therefore, since both theory and observation replicate this behaviour, it is not surprising that EAGLE’s simulations replicate it as well (e.g. Schaye et al. [2015]; Ferrero et al. [2017]).

As it can be seen in Figure 2.3, our own stellar TFR could be well approximated by a power law between the stellar mass of the galaxy and our chosen term  $V_{\max}$ .

$$Total : \quad \log_{10}(V_{\max}) \simeq 0.257 \cdot \log_{10}(M_*) - 0.405 \quad (3.1)$$

$$Disc : \quad \log_{10}(V_{\max}) \simeq 0.260 \cdot \log_{10}(M_*) - 0.444 \quad (3.2)$$

The slope, for the total number of stars, ( $s \simeq 0.257$ ) is found to be in agreement with measurements collected in Avila-Reese’s paper (Avila-Reese et al. [2008], Figure 1), because it presents reasonable values compared to the rest of the current studies (Stone et al. [2021]).

Besides, this result supports the statement made in Schaye’s paper (Schaye et al. [2015]) about this close relative of the Tully-Fisher relation (Schaye et al. [2015], Figure 12). It can therefore be concluded that the approximate value  $V_{\max}$  can be used as an appropriate reference value to describe the maximum rotational velocity of the galaxies simulated in Ref-L0050N0752. In addition, the results obtained in this work are trustworthy, and could be taken as representatives.

### 3.3 Distribution Functions

Distribution functions give probabilistic information about the property that is being studied. In this case, two studies have been performed, concerning the  $V_{\max}$  (Figure 2.4 first row) and the  $M_*$  of the galaxy (second row). Both properties are carried out in two parts. The first one (Figure 2.4 left column) consists of three superimposed histograms that compare the function distribution between different groups of massive galaxies according to their morphology: disc-type galaxies (our sample of interest), elliptical-type galaxies, and all of the above two. In this second part (right column), the histogram of disc-type galaxies is isolated, nonetheless the graph corresponding to the magnitude  $M_*$  has a curve associated with its GSMF.

This distribution of the images (Figure 2.4) allows an easy comparison between all of them.

- Top-left image of Figure 2.4: In this plot it is possible to examine the histograms of the  $V_{\max}$  of massive galaxies. There are three histograms in which both axes present a logarithmic scale. This distribution is typical in studies like this one, that have a large number of values, (e.g. the dataset based to display the values vary up to two orders of magnitude on each axis).

Firstly, it is possible to observe the distribution uniformity of the three galaxy set samples, the distribution is centred at a certain maximum and is followed by a decrease in the number of galaxies on either side. These descriptions added to the fact that all distributions do not have a forest of bars shape, suggest that the method used to obtain the distribution and data is effective and correct.

Secondly, one thing that can be verified in this graphic, The total galaxy distribution should be the sum of the other two distributions. In addition to verifying it numerically, it can be intuited because of two reasons: first, the total distribution is always greater than or equal to both distributions, and second, where the distribution of disc-type galaxies increases the total distribution differs more from the elliptical-type distribution.

It can also be seen that there are more elliptical galaxies than disc galaxies, that can be deduced because, although the galaxies are more or less equally distributed in terms of rotational velocity, almost all the bars of the elliptical galaxies are higher than those of the disc galaxies. Moreover, the fact that this is true, even though the dispersion of values of elliptical-type galaxies is larger than disc-type galaxies, reinforces the argument.

The distribution of elliptical-type galaxies has a larger range [1.73, 2.90] than the disc-type distribution [1.79, 2.58]. Furthermore, their maximum are not aligned, while most elliptical galaxies have a  $V_{\max}$  around  $10^2 \text{ km/s}$ , the disc galaxies are at  $10^{2.15} \text{ km/s}$ . Roughly speaking, they have similar values, but with respect to the range of distribution (it is a logarithm scale) there are considerable differences.

- bottom-left image of 2.4: Analogous to the previous discussion, but in this plot, the histograms refers to the study of the  $M_*$  for massive galaxies. These three bar graph also have logarithmic scales due to, again, the large range of data values.

The first thing that stands out is the lower limit presented by the image, this limit is due to the condition imposed in this thesis, stating that the sample studies of galaxies must be

massive. Therefore, the lower limit on the distribution of the  $M_*$  property is, in effect, at  $10^9 M_\odot$ .

Taking into account the above limitation and, comparing this image with the previous one, it can be seen that a description analogous to the previous set of histograms can be made: the bar graph of total galaxies is the sum of both histograms (the visual arguments expressed in the previous analysis are maintained), it is still observed that there is a greater amount of elliptical galaxies than disc galaxies, despite the fact that for this study there is a lower limit and the elliptical type galaxies present a greater range of distribution; Finally, the maximum of the distributions are not aligned, while the elliptical-type takes the maximum at the lower limit ( $10^9 M_{\text{dot}}$ , the disc galaxies are around  $10^{9.6} M_\odot$ ).

- Bottom-right image of 2.4: in addition to the histogram of  $M_*$  property of disc massive galaxies, the galaxy star function is drawn as a curve versus the density of galaxy number ( $\Delta \log_{10}(M_*) = 0.2$ ). It can be seen that, although it is a noisy curve, the behaviour of the curve can be intuited and when comparing it with the histogram it can be seen that it is a representative curve. Although at first it might seem that the noise of the curve is due to the lack of data from a larger number of galaxies, Table 2.1 shows, except at the upper limit of  $M_*$ , the great amount of data observed per bin. Schaye et al. [2015] find a worse case with the *Ref - L0025N0752* simulation (Figure 4) and find arguments that it is due to the size of the simulation.
- right row of Figure 2.4: Apart from everything described with the set of rotation velocity histograms, the distribution obtained in  $V_{\text{max}}$  for disc galaxies is related to the  $M_*$  distribution of the same galaxies. This statement can be intuited by visually comparing the two images, but strong arguments can also be given. Moreover, this similarity is well described by the TFR described and discussed above.

On the one hand, it can be observed that, although the stellar mass of the galaxies presents a lower boundary condition, no such boundary condition is observed in the maximum rotational velocity; nevertheless, it can be observed that the  $V_{\text{max}}$  distribution decays rapidly when it is on the order of  $10^{1.9} \text{km/s}$ , which is roughly equivalent to the stellar mass limit after using the fit for disc galaxies (3.2). The fact that even lower values of  $V_{\text{max}}$  than this limit are present may be due to the fact that the power law does not behave well at lower limits, so a more restrictive condition for massive galaxies could be considered for future work.

On the other hand, the rest of the distribution resembles the two plots quite well, both maxima are equivalent in terms of fit, and also show the same decreasing behaviour at the upper limits of both properties. With all this, we can again observe the clear relationship between both properties for the conditions imposed in this work.

Once these distributions have been compared, as well as the GSMF of massive disc galaxies, we also study the results of the GSMFs in 2.5. On the one hand, the code from Appendix A of McAlpine et al. [2016] is applied to obtain the two curves concerning the simulation chosen for this work, both for disc galaxies and for all massive galaxies (in Figure 2.5 are represented as orange and silver respectively). It can be seen how both correspond to their distributions presented in the bottom-right image of 2.4, that is why the GSMF of all galaxies decays later than that of the disc galaxies, because it takes into account the rest of the galaxies.

On the other hand, use is made of the literature to compare with two other GSMFs. First, the remaining curve also comes as GSMF from the EAGLE simulation Ref-L0100N1504 at  $z = 0.1$  (Schaye et al. [2015] Figure 4), it is the GSMF of all massive galaxies with a three-dimensional aperture of 30 kiloparsec proper (more details in section 3 of Schaye et al. [2015] and subsection 5.1.1)). Second, we show in Figure 2.5 some data points that come from Galaxy And Mass Assembly (GAMA) observations made by Baldry et al. [2012] (The set of data points are extracted as part of the results shown in Table 1). It can be seen how these last two illustrations present a similar compartmentalisation, which evidences the ability of EAGLE’s simulations to adequately represent GSMFs at low- $z$ .

It can be seen that the noise in the curves corresponding to the RefL0050N0752 simulation is similar, even though the curve representing the total set of massive galaxies has more galaxies per bin than the disc-only curves. In contrast, the Ref-L0100N1504 simulation has a much smoother curve.

## 3.4 Initial Conditions

As a last part of the work, it is proposed to find correlations between the galaxy rotation velocity distribution and the properties of the environments where the galaxies formed, in this case the over-density and the angular momentum of the region where the particles are located at redshift 20. This information is obtained from the code developed in this work, which is based on the code written during the internship.

### 3.4.1 The New Code

The modifications to the code allow: firstly, to use it for larger simulations than those tested in practice; secondly, to also extract angular momentum information; and thirdly, to have a more efficient program with a considerable increase in speed (up to 12 times faster). The latest version of the code is the one used to extract the simulation data Ref-L0050N0752.

The results provided in the subsection 2.3 (Figures 2.7, 2.8, 2.9, 2.10) prove that the code does indeed meet the objectives of the modifications specified in the methodology section (2). The improvements in the code can be seen in the time taken to obtain the overdensity of a galaxy. In the simulation Ref-L0012N0188, we went from 72 seconds per galaxy to just over 5 seconds per galaxy, 50 seconds for the whole sample. In the simulation that concerns this work it was not even possible to run with the initial code, with the current code it takes about 4 min per galaxy, which means that obtaining the over-density of all the galaxies concerned takes a little more than a day.

### 3.4.2 Over-density

With the code implemented, the properties obtained with this code can be studied. Firstly, identify the over-density ( $\Delta\rho$ ) of the regions where massive disc galaxies formed. The over-density distribution obtained for the EAGLE simulation Ref-L0050N0752 can be seen in Figure 2.7.



Looking at the histogram (Figure 2.7) it can be seen that it is a practically uniform distribution and does not have negative over-densities, i.e. regions with less matter density than the mean of the universe. Furthermore, it has a small distribution range, although it is on a logarithmic scale, a peak at 0.07 can be clearly seen. All this indicates that the study of the over-density can be held in good esteem, that in the first instance it does not present anomalies to question the result obtained.

The mean histogram is in 0.076, this equivalent to a mean over-density of 19% ( $\overline{\Delta\rho} \simeq 0.192$ ) with respect the mean density of the universe ( $\rho_0$ ), i.e. the mean density in the formation regions of massive disc galaxies is 1.19 times the mean density of the universe ( $\bar{\rho}_{disc} = 1.19\rho_0$ ).

With the over-density in the simulation verified, the next step, and the aim of this part, is researching whether there is a relationship between this property and the maximum rotational velocity. For this reason it is compared the  $V_{max}$  versus the  $\Delta\rho$  of the galaxies sample (Figure 2.9). it is possible to observe that there is a point cloud.

This result goes against the premise and could be taken as a first step to discard the idea that there is a correlation between the maximum rotation velocity and the over-density of the formation region of massive disc galaxies. Or at least, that their relationship is not so relevant and that there are other factors that have a greater impact on the dynamics of the galaxy.

Although our result does not shed much light, the right side of the figure is worth noting. If one maintains the idea that over-density and maximum rotational velocity are correlated, one would expect this relationship to be directly proportional, i.e. a region with a higher over-density would imply that the resulting galaxy would have a higher maximum rotational velocity. Looking at the right-hand side of the histogram, it can be seen that, although it is a point cloud, the regions with the highest  $\Delta\rho$  are those with the lowest  $V_{max}$ . This behaviour could be an argument for questioning the veracity of the data presented as over-density. This would allow us to maintain the idea that there is a correlation.

For this reason, two other density-related (Equation 2.6) cases are studied in order to present evidence that may reveal something of interest about the veracity of the data obtained that corresponds to over-density. These two properties are also properties of the region from which the massive disc galaxies formed: the radius of the selected volume as a region ( $R_s$ ) and the total mass inside the selected volume ( $M_s$ ). Both comparisons can be seen in 3.1.

As can be seen in the figure, both comparisons present a linear correlation, emphasising that the graphs compare  $\log_{10} - \log_{10}$ . In other words, they show some potential relationship of the following style:

$$\log_{10}(R_s) \simeq 1.78 \cdot \log_{10}(V_{max}) - 5.93 \quad (3.3)$$

$$\log_{10}(M_s) \simeq 2.62 \cdot \log_{10}(V_{max}) - 3.53 \quad (3.4)$$

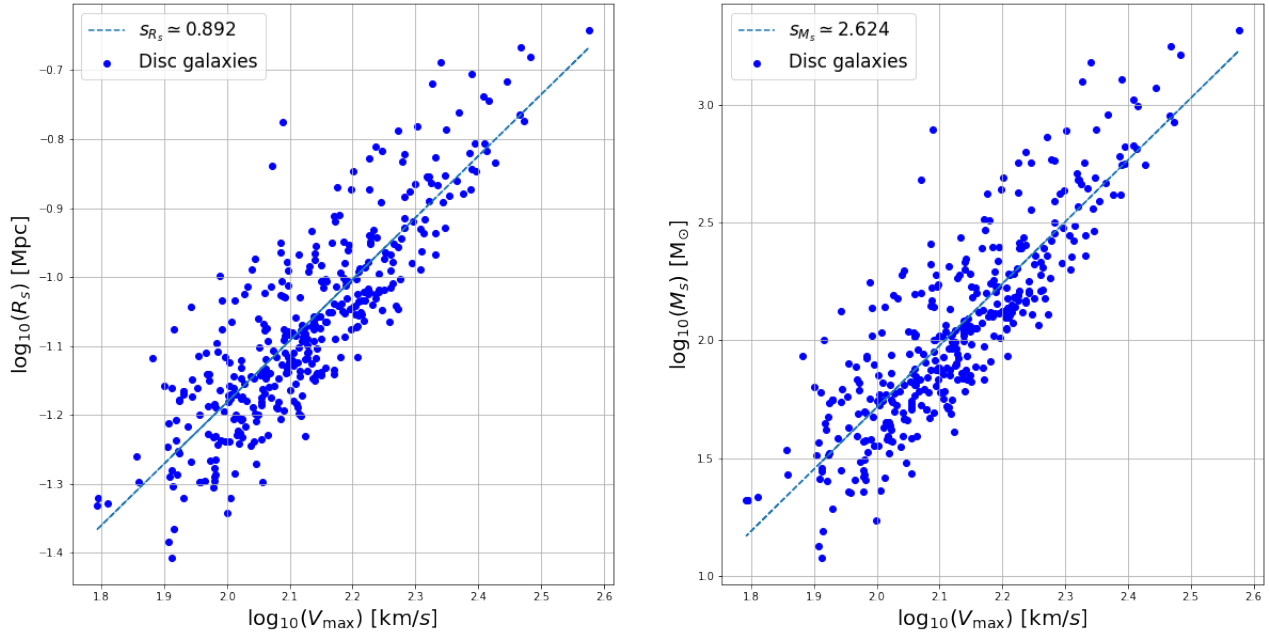


Figure 3.1:  $R_s$  (left image) and  $M_s$  (right image) versus  $V_{\max}$ . A study of two properties that compose the density, and hence the over-density, of the formation region of disc-type galaxies at  $z = 20$  with the maximum rotational velocity of these galaxies at  $z = 0$ . The relation of these results is proposed in order to shed some light on the veracity of the over-density.

On the one hand, the ratio of the maximum rotational velocity to the mass of the formation region has a simple explanation. The particles inside the region are practically the same as those that make up the galaxy. The mass of the stars in galaxies has a well-defined correlation with the mass of the halos that contain them. In addition, the maximum rotation velocity has a well-defined correlation with the mass in stars (the stellar TFR, Figure 2.3). So, the rotation speed has a correlation with the mass of the halo. This is quite logical, since the rotation speed depends on the distribution of all mass in the halo.

On the other hand, in the initial conditions of the simulated universe, the density of the entire universe is required to be in accordance with the mean density of the universe ( $\rho_0$ ), in order to replicate the Cosmological Principle on which cosmological models such as  $\Lambda$ CDM are based. This correlation would explain why  $R_s$  and  $V_{\max}$  also correlate.

Understanding how these properties correlate with each other (Figure 3.2), with the  $V_{\max}$  (Equation 3.3 and Figure 3.1) and with  $\rho$  (Equation 2.6) explains why over-density as defined does not correlate in use (since it is given by a non-linear combination of both properties). Surely with a larger number, for example with a larger simulation and with a larger sample of galaxies, the relationship between over-density and maximum galaxy rotation velocity can be more easily discerned.

From this point on, one could consider using a different definition of density than the one proposed in this thesis. One way to take is to change the already questioned definition that has been used as the radius of the volume of the galaxy formation region. One could study the density inside

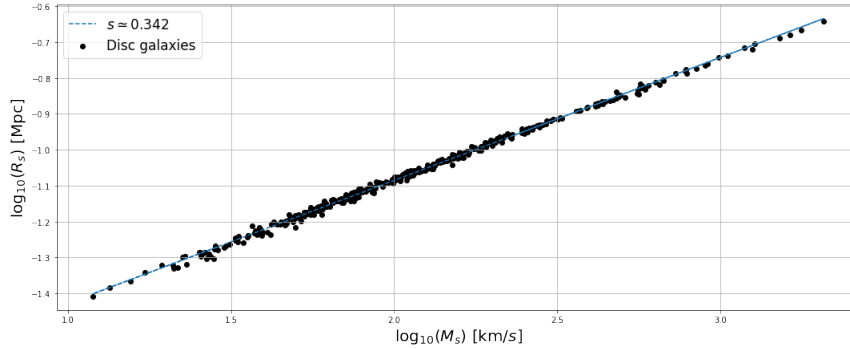


Figure 3.2:  $M_s$  versus  $R_s$  of massive disc galaxies in EAGLE's Ref-L0050N0752 simulation at  $z = 20$ .

The density distribution must not be exactly uniform throughout the universe in order to give rise to cosmological structures, due to the gravitational interaction originating from over-density regions such as those studied in this work. At  $z = 20$ , the mass of a region is strongly correlated with the volume of that region, which implies that the over-densities of the regions will not be very far from the average density of the universe (as tested in this section).

a sphere with a centre equal to the centre of mass but with a certain radius that is the same for all galaxies ( $R_{s,j} = R_s \forall j$ ); with this new definition of density, and therefore of over-density, one can study galaxy size indifferently, since smaller galaxies have a faster formation cycle than larger galaxies.

### 3.4.3 Angular Momentum

The discussion of the angular momentum data is similar to the first part of the discussion of the over-density data. Again the data histogram shows a uniform distribution of angular momentum, although there is a disk galaxy in the sample set that has an angular momentum whose modulus is less than unity (with unity being  $M_\odot \text{Mpc km s}^{-1}$  (left limit of Figure 2.8)).

The aim of this part of the work is to examine whether there is any relationship between the angular momentum and the maximum rotational velocity, The Figure 2.10 visually presents the behaviour of one versus the other. It can be seen that the result obtained is a point cloud, which implies that there is apparently no correlation between the two magnitudes.

Unlike the over-density, the fact that the angular momentum has no clear relationship (at least not visible in this thesis) comes as a surprise. This is because it is assumed that the total angular momentum of a galaxy is conserved, so if the mass of the galaxy tends not to vary much and the radius of the galaxy decreases due to gravitational interaction, the rotational velocity should increase. The most interesting answer that can be drawn from this study is that, as with the over-density, the definition of angular momentum in the region of galaxy formation is not ideal for finding this relation.

Another cause may be that the angular momentum term is being considered too early to correctly express the expected behaviour, remembering that it is studied from the region where the galaxy will form. The tidal torque of the total set of particles that will later form the galaxy is a property from

which it is theorised that the angular momentum of galaxies can emerge (see for more details Barnes and Efstathiou [1987]; and for observational testing Porciani et al. [2002]), which would provide information on the formation of disc galaxies (see Fall and Efstathiou [1980]), thus explaining why the angular momentum obtained is not conserved in the proto-galaxy regions (see White [1984]). The disadvantage of this property lies in its complexity: it is a tensor from which each of the components would have to be obtained, which increases the computational time required per number of galaxies, which would make the computational problem when wanting to use a larger and higher quality simulation a real challenge to work on.

### 3.5 Conclusions

This work has enabled an elaborate and comprehensive study of the rotational velocity of massive disc galaxies. It has allowed us to understand the importance of simulations in the field of research and how they complement both theory and observation. The following conclusions can be drawn from this work:

The use of the Ref-L0050N0752 simulation seems to be a good balance between a large enough sample size for its results to show some relevance but without excessive computational complexity that would make it impossible to obtain the necessary data. 385 galaxies is a significant sample to carry out the statistical studies shown in this publication, A smaller simulation and the number of optimal galaxies would have been too low, a larger simulation and the computational demands would not have allowed the data to be obtained. Furthermore, it is a sample that is under the conditions necessary to carry out this investigation. They are all massive and are the central galaxy of their own halo, which means that the study does not have to survey the order of perturbation associated with gravitational interaction with other bodies in the same halo.

The TFR obtained in this work is satisfactory, it is a relationship that is measured observationally and the fact that it is reproduced in the simulations is a good step forward. With this section it is possible to affirm that the  $V_{\max}$  represents in a good way the maximum rotation velocity of galaxies even if it is obtained from a theoretical approximation, which allows us to continue with the study.

The distribution functions of the different properties proposed in this study show a good performance. All the histograms studied in this section are free of anomalies and strange data that make us question the reliability of the distributions obtained. The relation for the set of disc galaxies between the distribution of their stellar mass and their maximum rotation velocity reconfirms what was said about their TFR. The fact that this relation can be based on the relation obtained in the TFR reinforces the first part of this study as well as the reliability of the results.

The GSMFs used in this thesis show the potential of their use. These functions allow to relate and compare between studies of different modalities (analytical, numerical, observational) for the same field. The results obtained in our sample of galaxies show a good performance against observational data and other simulations. This good result reinforces the good view we already have on the distribution function obtained for the stellar mass of massive disc galaxies.

With all of the above, we conclude that the distribution of the  $V_{\max}$  function for massive disc galaxies obtained in this work is sufficiently well supported to claim that it is a good representation of what we were looking for in this work.

The code is the most extensive section of the work, from its creation until the moment of the delivery of the memory it is in constant change and improvement. It can be stated that it is a functional code and that it can be used optimally for the work employed in this thesis, but it may need more changes and improvements if it is to be used for larger simulations or for tests with greater stress. The class *Data\_snapnum* is the epicentre of the improvements, it is the main class of the whole program, this is due to the fact that all the calculations involve extracting the data from the files, a function that this class performs. It is where most relevant modifications are made with respect to the original code (changes mentioned in 2.3). The rest of the functions (*Galaxy\_to\_past*, *Overrho* and *AngularMoment*) present a non-optimised behaviour with possible bottlenecks that affect the performance of the code.

The improvement and optimisation of the code becomes crucial because it is not a linear computational problem, as the increase in complexity from one dimension to another not only has to take into account the increase in the number of galaxies, but also the quality of each one of them, i.e. the number of particles in each galaxy. A possible future advance in this line of research could consist of improving these functions, with special mention to the *in1d* function of the *Numpy* library, with the aim of eliminating the possible bottlenecks that arise in the code and thus being able to carry out simulations of a larger size and with a better quality.

No evidence is found for a robust connection between the two properties of the galaxy-forming regions of the samples selected for study, density and angular momentum, and the maximum rotational velocity of the galaxies.

In one hand, the over-density case. Despite the poor results, the background to why over-density did not correlate is probably understood. The correlations between  $M_s$  and  $R_s$  with  $V_{\max}$  show the non-trivial way in which the latter is linked to  $V_{\max}$ . In addition, new ways of defining the over-density, such as an  $R_s$  equal to all galaxies, are proposed for future work.

On the other hand, the angular momentum background has not been so easy to discern. And it is from here that it is proposed to continue future work. Despite the possible computational challenges, the study of the tidal torque of the formation region may shed light on this issue.

Despite the negative results, the formal response of this work is to keep the hypothesis open that both over-density and angular momentum in the formation regions, especially the latter, are correlated with the maximum rotational velocity. Alternative definitions and properties relative to those of this work will be investigated in order to continue this part of the work.

# Bibliography

- 168–173, mar 1929. doi: 10.1073/pnas.15.3.168. URL <https://doi.org/10.1073%2Fpnas.15.3.168>.
- Eleonora Di Valentino, Olga Mena, Supriya Pan, Luca Visinelli, Weiqiang Yang, Alessandro Melchiorri, David F. Mota, Adam G. Riess, and Joseph Silk. In the Realm of the Hubble tension - a Review of Solutions. *arXiv e-prints*, art. arXiv:2103.01183, March 2021.
- J.J. Condon and A.M. Matthews.  $\Lambda$ CDM Cosmology for Astronomers. *passp*, 130(989):073001, July 2018. doi: 10.1088/1538-3873/aac1b2.
- Planck Collaboration, P.A.R. Ade, N. Aghanim, C. Armitage-Caplan, Arnaud, et al. Planck 2013 results. XVI. Cosmological parameters. *aap*, 571:A16, November 2014. doi: 10.1051/0004-6361/201321591.
- Albert Einstein. The Field Equations of Gravitation. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys. )*, 1915:844–847, 1915.
- S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S, et al. Measurements of  $\omega$  and  $\lambda$  from 42 high-redshift supernovae. *apj*, 517(2):565–586, jun 1999. doi: 10.1086/307221. URL <https://doi.org/10.1086/307221>.
- The EAGLE team. The EAGLE simulations of galaxy formation: Public release of particle data. *arXiv e-prints*, art. arXiv:1706.09899, June 2017.
- Joop Schaye, Robert A. Crain, Richard G. Bower, Michelle Furlong, Matthieu Schaller, Tom Theuns, Claudio Dalla Vecchia, et al. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *mnras*, 446(1):521–554, January 2015. doi: 10.1093/mnras/stu2058.
- Houjun Mo, Frank van den Bosch, and Simon White. Galaxy formation and evolution. In *Galaxy Formation and Evolution*, pages 0–842. Cambridge, mar 2013. doi: 10.1017/cbo9780511807244.023. URL <https://oadoi.org/10.1017/cbo9780511807244.023>.
- E. Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3): 168–173, mar 1929. doi: 10.1073/pnas.15.3.168. URL <https://doi.org/10.1073%2Fpnas.15.3.168>.
- S. McAlpine, J. C. Helly, M. Schaller, J. W. Trayford, Y. Qu, M. Furlong, R. G. Bower, R. A. Crain, J. Schaye, T. Theuns, C. Dalla Vecchia, C. S. Frenk, I. G. McCarthy, A. Jenkins, Y. Rosas-Guevara, S. D. M. White, M. Baes, P. Camps, and G. Lemson. The EAGLE simulations of galaxy formation: Public release of halo and galaxy catalogues. *Astronomy and Computing*, 15:72–89, April 2016. doi: 10.1016/j.ascom.2016.02.004.
- Robert A. Crain, Joop Schaye, Richard G. Bower, Michelle Furlong, Matthieu Schaller, Tom Theuns, Claudio Dalla Vecchia, et al. The EAGLE simulations of galaxy formation: calibration of subgrid physics and model variations. *mnras*, 450(2):1937–1961, June 2015. doi: 10.1093/mnras/stv725.
- Volker Springel. The cosmological simulation code gadget-2. *mnras*, 364(4):1105–1134, December 2005. doi: 10.1111/j.1365-2966.2005.09655.x.
- M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White. The evolution of large-scale structure in a universe dominated by cold dark matter. *apj*, 292: 371–394, May 1985. doi: 10.1086/163168.
- Volker Springel, Simon D. M. White, Giuseppe Torren, and Guinevere Kauffmann. Populating a cluster of galaxies - I. Results at  $z=0$ . *mnras*, 328(3):726–750, December 2001. doi: 10.1046/j.1365-8711.2001.04912.x.
- K. Dolag, S. Borgani, G. Murante, and V. Springel. Substructures in hydrodynamical cluster simulations. *mnras*, 399(2):497–514, October 2009. doi: 10.1111/j.1365-2966.2009.15034.x.
- Mark T Graham, Michele Cappellari, Hongyu Li, Shude Mao, Matthew A Bershady, et al. SDSS-IV MaNGA: stellar angular momentum of about 2300 galaxies: unveiling the bimodality of massive galaxy properties. *Monthly Notices of the Royal*

- Astronomical Society*, 477(4):4711–4737, 03 2018. ISSN 0035-8711. doi: 10.1093/mnras/sty504. URL <https://doi.org/10.1093/mnras/sty504>.
- Michael R. Blanton, David W. Hogg, Neta A. Bahcall, Ivan K. Baldry, J. Brinkmann, et al. The Broadband Optical Properties of Galaxies with Redshifts  $0.02 < z < 0.22$ . *apj*, 594(1):186–207, September 2003. doi: 10.1086/375528.
- S. P. Driver, P. D. Allen, Alister W. Graham, E. Cameron, J. Liske, et al. The Millennium Galaxy Catalogue: morphological classification and bimodality in the colour-concentration plane. *mnras*, 368(1):414–434, May 2006. doi: 10.1111/j.1365-2966.2006.10126.x.
- R.B. Tully and J.R. Fisher. A new method of determining distance to galaxies. *aap*, 500:105–117, February 1977.
- A. Dressler. Galaxy morphology in rich clusters: implications for the formation and evolution of galaxies. *apj*, 236:351–365, March 1980. doi: 10.1086/157753.
- I.K. Baldry, M.L. Balogh, R.G. Bower, K. Glazebrook, R.C. Nichol, and others. Galaxy bimodality versus stellar mass and environment. *mnras*, 373(2):469–483, December 2006. doi: 10.1111/j.1365-2966.2006.11081.x.
- Lee S. Kelvin, Simon P. Driver, Aaron S.G. Robotham, Edward N. Taylor, Alister W. Graham, et al. Galaxy And Mass Assembly (GAMA): stellar mass functions by Hubble type. *mnras*, 444(2):1647–1659, October 2014. doi: 10.1093/mnras/stu1507.
- Jr. Kennicutt, R.C. The rate of star formation in normal disk galaxies. *apj*, 272:54–67, September 1983. doi: 10.1086/161261.
- V. Avila-Reese, J. Zavala, C. Firmani, and H.M. Hernández-Toledo. On the Baryonic, Stellar, and Luminous Scaling Relations of Disk Galaxies. *aj*, 136(3):1340–1360, September 2008. doi: 10.1088/0004-6256/136/3/1340.
- Camila A. Correa, Joop Schaye, Bart Clauwens, Richard G. Bower, Robert A. Crain, Matthieu Schaller, Tom Theuns, and Adrien C.R. Thob. The relation between galaxy morphology and colour in the EAGLE simulation. *mnras*, 472(1):L45–L49, November 2017. doi: 10.1093/mnrasl/slx133.
- Adrien C.R. Thob, Robert A. Crain, Ian G. McCarthy, Matthieu Schaller, et al. The relationship between the morphology and kinematics of galaxies and its dependence on dark matter halo structure in EAGLE. *mnras*, 485(1):972–987, May 2019. doi: 10.1093/mnras/stz448.
- K.C. Freeman. On the Disks of Spiral and S0 Galaxies. *apj*, 160:811, June 1970. doi: 10.1086/150474.
- Vera C. Rubin and Jr. Ford, W. Kent. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *apj*, 159:379, February 1970. doi: 10.1086/150317.
- Yoshiaki Sofue and Vera Rubin. Rotation Curves of Spiral Galaxies. *araa*, 39:137–174, January 2001. doi: 10.1146/annurev.astro.39.1.137.
- K. Freese. Review of Observational Evidence for Dark Matter in the Universe and in upcoming searches for Dark Stars. In E. Pécontal, T. Buchert, Ph. di Stefano, and Y. Copin, editors, *EAS Publications Series*, volume 36 of *EAS Publications Series*, pages 113–126, January 2009. doi: 10.1051/eas/0936016.
- M. Pierce. *Tully-Fisher Relation*, page 2622. Murdin, P., 2000. doi: 10.1888/0333750888/2622.
- Connor Stone, Stéphane Courteau, and Nikhil Arora. The Intrinsic Scatter of Galaxy Scaling Relations. *apj*, 912(1):41, May 2021. doi: 10.3847/1538-4357/abebe4.
- V. Avila-Reese and C. Firmani. The evolution of disk galaxies. In Jorge Cantó and Luis F. Rodríguez, editors, *Revista Mexicana de Astronomía y Astrofísica Conference Series*, volume 10 of *Revista Mexicana de Astronomía y Astrofísica Conference Series*, pages 97–100, March 2001.
- Ismael Ferrero, Julio F. Navarro, Mario G. Abadi, Laura V. Sales, Richard G. Bower, Robert A. Crain, Carlos S. Frenk, Matthieu Schaller, Joop Schaye, and Tom Theuns. Size matters: abundance matching, galaxy sizes, and the Tully-Fisher relation in EAGLE. *mnras*, 464(4):4736–4746, February 2017. doi: 10.1093/mnras/stw2691.

- Anna K. Weigel, Kevin Schawinski, and Claudio Bruderer. Stellar mass functions: methods, systematics and results for the local Universe. *mnras*, 459(2):2150–2187, June 2016. doi: 10.1093/mnras/stw756.
- Maarten Schmidt. Space Distribution and Luminosity Functions of Quasi-Stellar Radio Sources. *apj*, 151:393, February 1968. doi: 10.1086/149446.
- Cheng Li and Simon D. M. White. The distribution of stellar mass in the low-redshift Universe. *mnras*, 398(4):2177–2187, October 2009. doi: 10.1111/j.1365-2966.2009.15268.x.
- I. K. Baldry, S. P. Driver, J. Loveday, E. N. Taylor, L. S. Kelvin, J. Liske, P. Norberg, A. S. G. Robotham, S. Brough, A. M. Hopkins, S. P. Bamford, J. A. Peacock, J. Bland-Hawthorn, C. J. Conselice, S. M. Croom, D. H. Jones, H. R. Parkinson, C. C. Popescu, M. Prescott, R. G. Sharp, and R. J. Tuffs. Galaxy And Mass Assembly (GAMA): the galaxy stellar mass function at  $z \leq 0.06$ . *mnras*, 421(1):621–634, March 2012. doi: 10.1111/j.1365-2966.2012.20340.x.
- P. Bonfini, A. Zezas, M. L. N. Ashby, S. P. Willner, A. Maragkoudakis, K. Kouroumpatzakis, P. H. Sell, and K. Kovelakos. The star formation reference survey - IV. Stellar mass distribution of local star-forming galaxies. *mnras*, 504(3):3831–3861, July 2021. doi: 10.1093/mnras/stab777.
- P. Schechter. An analytic expression for the luminosity function for galaxies. *apj*, 203:297–306, January 1976. doi: 10.1086/154079.
- H. J. Mo, Shude Mao, and Simon D. M. White. The formation of galactic discs. *mnras*, 295(2):319–336, April 1998. doi: 10.1046/j.1365-8711.1998.01227.x.
- Oleg Y. Gnedin, David H. Weinberg, James Pizagno, Francisco Prada, and Hans-Walter Rix. Dark Matter Halos of Disk Galaxies: Constraints from the Tully-Fisher Relation. *apj*, 671(2):1115–1134, December 2007. doi: 10.1086/523256.
- Stéphane Courteau, Aaron A. Dutton, Frank C. van den Bosch, Lauren A. MacArthur, Avishai Dekel, Daniel H. McIntosh, and Daniel A. Dale. Scaling Relations of Spiral Galaxies. *apj*, 671(1):203–225, December 2007. doi: 10.1086/522193.
- Eric F. Bell, Carlton M. Baugh, Shaun Cole, Carlos S. Frenk, and Cedric G. Lacey. The properties of spiral galaxies: confronting hierarchical galaxy formation models with observations. *mnras*, 343(2):367–384, August 2003. doi: 10.1046/j.1365-8711.2003.06673.x.
- Sebastian Trujillo-Gomez, Anatoly Klypin, Joel Primack, and Aaron J. Romanowsky. Galaxies in  $\Lambda$ CDM with Halo Abundance Matching: Luminosity-Velocity Relation, Baryonic Mass-Velocity Relation, Velocity Function, and Clustering. *apj*, 742(1):16, November 2011. doi: 10.1088/0004-637X/742/1/16.
- Joshua Barnes and George Efstathiou. Angular Momentum from Tidal Torques. *apj*, 319:575, August 1987. doi: 10.1086/165480.
- Cristiano Porciani, Avishai Dekel, and Yehuda Hoffman. Testing tidal-torque theory - I. Spin amplitude and direction. *mnras*, 332(2):325–338, May 2002. doi: 10.1046/j.1365-8711.2002.05305.x.
- S. M. Fall and G. Efstathiou. Formation and rotation of disc galaxies with haloes. *mnras*, 193:189–206, October 1980. doi: 10.1093/mnras/193.2.189.
- S. D. M. White. Angular momentum growth in protogalaxies. *apj*, 286:38–41, November 1984. doi: 10.1086/162573.



# Appendix A

## Code to Obtain the Dataset

### Introduction

```
# -*- coding: utf-8 -*-
"""
    TFG_code.py — version 4.14.0 —
"""

# Libraries
import os
import h5py
import gc
import numpy as np
import eagleSqlTools as eagle

# List of EAGLE's simulations.
list_sim = ["RefL0012N0188", "RefL0025N0376",
            "RefL0050N0752", "RecalL0025N0752"]

sim = list_sim[0]      # Simulation selected
```

### Data\_snapnum Class

```
"""
    # Class Data_snapnum
"""

# Class to get the data from a snapshot of a simulation.
class Data_snapnum():
    def __init__(self, simulation, snapnum, path,
                 user_name="<user-name>", password="<password>"):
        """
            To load the class is requested:
            simulation : The EAGLE's simulation
            snapnum    : The number related to the snapshot of this simulation.
            Path       : Path where the .hdf5 file with the snapshot information.
            user_name, password: The username and password of an EAGLE account.
        """

        # The input data is stored in the object.
        self.user = {'username': user_name, 'password': password}
        self.sim = simulation
        self.snapnum = snapnum
        self.path = path
        self.subpath = "snap_"+self.path[len("/%s/snapshot"%(self.sim)):]
        self.nfiles = len(os.listdir("%s/%s"%(path)))

        # Certain interesting constants are loaded.
        with h5py.File('%s/%s.%i.hdf5'%(self.path, self.subpath, 0), 'r') as f:
            """
                # Scale factor
                # Hubble constan: 0.667
                # DarkMatter mass (M_*)
            """
            self.a = f['Header'].attrs.get('Time')
            self.h = f['Header'].attrs.get('HubbleParam')
```

```

self.mDM = f['Header'].attrs.get('MassTable')[1]*self.h**[-1]

## Other Certain interesting constants are loaded (rho_c)
## critical density (g/cm^3)
## matter density (g/cm^3)
## average matter density at this snapshot.
self.rho_c = 30000/((3.086e+19)**2*6.6743e-8*np.pi)*(self.h)**2
Omega_matter = f['Header'].attrs.get('Omega0')
self.rho_matter = self.rho_c*Omega_matter*self.a**(-3)

aexp = f['_PartType%i/%s'%(0, 'Coordinates')].attrs.get('aexp-scale-exponent')
hexp = f['_PartType%i/%s'%(0, 'Coordinates')].attrs.get('h-scale-exponent')

## Size of the universe (Mpc)
self.boxsize = f['Header'].attrs.get('BoxSize')*self.a**(aexp)*self.h**(hexp)

del(aexp, hexp, Omega_matter)
gc.collect()

## load the Galaxies catalogue. (Snapshot 0 hasn't)
if self.snapnum != 0:
    self.load_catalogue()

## Function to load certain galaxy data.
def load_catalogue(self):
    ##
    ## Function that allows to load the different galaxies with certain
    ## properties such as their GroupNumber and SubGroupNumber.
    ##
    ## This is valid for all that is different from snapnum 0, because it is
    ## from that snapnum that the GroupNumber and SubGroupNumber that the
    ## GroupNumber and SubGroupNumber are registered.
    ##
    ## Query information:
    ## Snapnum: Snapshot's number
    ## Galaxy ID: Number for each galaxy at snapnum number
    ## GroupNumber: Halo's number
    ## SubGroupNumber: Galaxy's number
    ## KappaCoRot: Disc parameter (if KappaCoRot > 0.4 -> Galaxy Disc-like)
    ##

    ## Connection with the database
    con = eagle.connect(user = self.user['username'],
                        password = self.user['password'])
    ## Query with the conditions for massive disc galaxies
    query = "SELECT_\
    -----MK.GalaxyID,_\
    -----SH.GroupNumber,_\
    -----SH.SubGroupNumber,_\
    -----SH.Vmax,_\
    -----SH.MassType_Star,_\
    -----SH.Mass,_\
    -----SH.GasSpin_z,_\
    -----MK.KappaCoRot_\
    -----FROM_\
    -----%s_SubHalo_AS_SH,_\
    -----%s_MorphoKinem_AS_MK_\
    -----WHERE_\
    -----SH.GalaxyID=_MK.GalaxyID_AND_\
    -----SH.MassType_Star>=1E09_AND_\
    -----MK.KappaCoRot>=0.4_AND_\
    -----SH.SubGroupNumber=_0_AND_\
    -----SH.SnapNum=_%i_\
    -----ORDER_BY_\
    -----SH.GalaxyID"%(self.sim, self.sim, self.snapnum)

    ## Our Catalogue
    self.catalogue = eagle.execute_query(con, query)

    del(con, query)
    gc.collect()

## Function to read the data
def read_dataset(self, itype, att):
    ##
    ## itype is the type of the particle:
    ## PartType0 -> Gas particle data
    ## PartType1 -> Dark matter particle data
    ## PartType4 -> Star particle data
    ## PartType5 -> Black hole particle data
    ## att is the attribute which is looking for.
    ##

    ## File upload to know the number particles.
    with h5py.File('%s/%s.%i.hdf5'%(self.path, self.subpath, 0), 'r') as f:
        ## Particles' number.

```

```

n_particles = f['Header'].attrs.get('NumPart.Total')[itype]

# Output array for Coordinates and Velocity cases.
if att == 'Coordinates' or att == 'Velocity':
    data = np.ones((n_particles, 3))

# Output array rest of cases
else:
    data = np.ones(n_particles)

del(n_particles)
gc.collect

# DarkMatter Mass case
if itype==1 and att=='Mass':
    data *= self.m_DM
    data.dtype = [('Mass', data.dtype)]
    return data

# The rest
else:
    """
    # There are failures when the specified itype is not present the attribute,
    so a it is used a try-except estructure
    """
    try:
        count = 0
        # Loop over each file and extract the data.
        for i in range(self.nfiles):

            f = h5py.File('%s/%s.%i.hdf5'%(self.path, self.subpath, i), 'r')
            tmp = f['PartType%i/%s'%(itype, att)][...]
            data[count:count+len(tmp)] = tmp
            count += len(tmp)

            """
            In the paper (arXiv: 1706.09899, part 4.1) repeat these calcs for each
            nfile (i.e. i value) but they didn't save in any variable, they overwrite all
            these variables, so it's fast if we do only in the last one, or first or whatever.

            I choose the last one.
            """
            aexp = f['PartType%i/%s'%(itype, att)].attrs.get('aexp-scale-exponent')
            hexp = f['PartType%i/%s'%(itype, att)].attrs.get('h-scale-exponent')

            f.close()
            del(f, tmp, count, i)
            gc.collect()

            # convert comovil to physical units and eliminate h multiplications
            if att != 'ParticleIDs' and data.dtype != np.int32 and data.dtype != np.int64:
                data = np.multiply(data, self.a**aexp * self.h**hexp, dtype='f8')

            del(aexp, hexp, tmp)
            gc.collect()

# In case there are no particles of that itype in this snapshot
except KeyError:
    if att == 'Coordinates' or att == 'Velocity':
        data = np.ones(3)*np.nan
    else:
        data = np.array([np.nan])

finally:
    if att == 'Coordinates' or att == 'Velocity':
        data.dtype = [(att+'_%i'%(i), data.dtype) for i in [0, 1, 2]]

    elif att == 'ParticleIDs':
        data.dtype = [('ParticleIDs', '<u8')]

    else:
        data.dtype = [(att, data.dtype)]

    return data

def periodicity(self, array, point, center=False):
    """
    A function that allows particles to be centred around the coordinates
    of a point or a particle, based on the periodicity property of the universe.
    """

    # Particle as a point of reference:
    if point.dtype == np.dtype([('ParticleIDs', '<u8'), ('Coordinates-0', '<f8'),
                                ('Coordinates-1', '<f8'), ('Coordinates-2', '<f8'),
                                ('Mass', '<f8'), ('itype', 'i1')]):
        for i in [0, 1, 2]:
            i_point = point['Coordinates_%i'%(i)]

```

```

# Particle as a reference point
array['Coordinates_%i'%i] == i_point

# For all particle beyond L/2...
mask = array['Coordinates_%i'%i] > self.boxsize/2
# ... is placed on the other side
array['Coordinates_%i'%i] == mask.astype(np.int)*self.boxsize
del(mask)
gc.collect()
# For all particle beyond -L/2...
mask = array['Coordinates_%i'%i] < -self.boxsize/2
# ... is placed on the other side
mask = mask.astype(np.int)*self.boxsize
array['Coordinates_%i'%i] += mask

# IF center = False -> the coordinates are retrieved
if not(center):
    array['Coordinates_%i'%i] += i_point

del(mask, i_point)
gc.collect()

del(i, point)
gc.collect()
return array

# Point as reference point case (The rest is the same)
else:
    for i in [0, 1, 2]:
        array['Coordinates_%i'%i] == point[i]

        mask = array['Coordinates_%i'%i] > self.boxsize/2
        array['Coordinates_%i'%i] == mask.astype(np.int)*self.boxsize

        del(mask)
        gc.collect()

        mask = array['Coordinates_%i'%i] < -self.boxsize/2
        mask = mask.astype(np.int)*self.boxsize
        array['Coordinates_%i'%i] += mask

        if not(center):
            array['Coordinates_%i'%i] += point[i]

    del(mask, i)
    gc.collect()

    return array

# Function to obtain the particle data of a certain sector
def particles_prop(self, att=None, itype=None, gn=None, sgn=None):
    """
    Possibility to upload a quantity of data directly. Depends on the need.
    att -> specific attribute
    itype -> specific particle-type
    gn -> specific GroupNumber
    sgn -> specific SubGroupNumber
    """

    if itype == None and att == None:
        """
        For all particles: ID, coordinates, mass and itype attribute
        """
        with h5py.File('%s/%s.%i.hdf5'%(snap_0.path, snap_0.subpath, 0), 'r') as f:
            n_particles = f['Header'].attrs.get('NumPart_Total')
            DF = np.ones(np.sum(n_particles),
                        dtype=np.dtype([('ParticleIDs', 'u8'), ('Coordinates_0', '<f8'), \
                                      ('Coordinates_1', '<f8'), ('Coordinates_2', '<f8'), \
                                      ('Mass', '<f8'), ('itype', 'i1')]))

        count = 0
        for itype, i in zip([0, 1, 4, 5], [0, 1, 2, 3]):

            if n_particles[i] == 0:
                continue

            for att in ['ParticleIDs', 'Coordinates', 'Mass', 'itype']:
                if att == 'Coordinates':
                    data = snap_0.read_dataset(itype, att)
                    DF['Coordinates_0'][count:count+n_particles[i]] = data['Coordinates_0'][..., 0]
                    DF['Coordinates_1'][count:count+n_particles[i]] = data['Coordinates_1'][..., 0]
                    DF['Coordinates_2'][count:count+n_particles[i]] = data['Coordinates_2'][..., 0]
                    del(data)
                    gc.collect()

                elif att == 'itype':
                    DF['itype'][count:count+n_particles[i]] *= itype

```

```

    else:
        DF[att][count:count+n_particles[i]] = snap_0.read_dataset(itype= itype ,att= att)

        count += n_particles[i]

del(count, n_particles, att, itype, i)
gc.collect()

elif itype == None:
    """
    For all particles: the attribute selected
    """
    dicc_dtype = { 'ParticleIDs':[( 'ParticleIDs', 'u8')],
                  'Coordinates':[( 'Coordinates_0', '<f8'),
                                  ( 'Coordinates_1', '<f8'),
                                  ( 'Coordinates_2', '<f8')],
                  'Mass': [( 'Mass', '<f8')],
                  'itype': [( 'itype', 'i1')] }

with h5py.File('%s/%s.%i.hdf5'%(snap_0.path, snap_0.subpath, 0), 'r') as f:
    n_particles = f['Header'].attrs.get('NumPart_Total')
    DF = np.ones(np.sum(n_particles), dtype=np.dtype(dicc_dtype[att]))

count = 0
for itype, i in zip([0, 1, 4, 5], [0, 1, 2, 3]):

    if n_particles[i] == 0:
        continue

    if att == 'Coordinates':
        data = snap_0.read_dataset(itype, att)
        DF['Coordinates_0'][count:count+n_particles[i]] = data['Coordinates_0'][..., 0]
        DF['Coordinates_1'][count:count+n_particles[i]] = data['Coordinates_1'][..., 0]
        DF['Coordinates_2'][count:count+n_particles[i]] = data['Coordinates_2'][..., 0]
        del(data)
        gc.collect()

    elif att == 'itype':
        DF['itype'][count:count+n_particles[i]] *= itype

    else:
        DF[att][count:count+n_particles[i]] = snap_0.read_dataset(itype= itype ,att= att)

        count += n_particles[i]

del(count, n_particles, itype, i)
gc.collect()

else:
    """
    For the itype particles selected: the selected attribute
    """
with h5py.File('%s/%s.%i.hdf5'%(snap_0.path, snap_0.subpath, 0), 'r') as f:
    n_particles = f['Header'].attrs.get('NumPart_Total')[itype]
    DF = np.ones(n_particles, dtype=np.dtype([( 'ParticleIDs', 'u8'), ( 'Coordinates_0', '<f8'), \
                                              ( 'Coordinates_1', '<f8'), ( 'Coordinates_2', '<f8'), \
                                              ( 'Mass', '<f8'), ( 'itype', 'i1')]))

count = 0
if n_particles[i] == 0:
    pass
else:
    for att in ['ParticleIDs', 'Coordinates', 'Mass', 'itype']:
        if att == 'Coordinates':
            data = snap_0.read_dataset(itype, att)
            DF['Coordinates_0'][count:count+n_particles[i]] = data['Coordinates_0'][..., 0]
            DF['Coordinates_1'][count:count+n_particles[i]] = data['Coordinates_1'][..., 0]
            DF['Coordinates_2'][count:count+n_particles[i]] = data['Coordinates_2'][..., 0]
            del(data)
            gc.collect()

        elif att == 'itype':
            DF['itype'][count:count+n_particles[i]] *= itype

        else:
            DF[att][count:count+n_particles[i]] = snap_0.read_dataset(itype= itype ,att= att)

        count += n_particles[i]

del(count, n_particles, att)
gc.collect()

```

```

if gn != None and sgn != None and self.snapnum != 0:
    """
    the data obtained is specified for a gn and sgn selected.
    NOTE: The zeroth snapshot does not GroupNumber nor SubGroupNumber
    """
    gns = self.read_dataset(itype, 'GroupNumber')['GroupNumber']
    sgn = self.read_dataset(itype, 'SubGroupNumber')['SubGroupNumber']
    mask = np.logical_and(gns == gn, sgn == sgn) # Mask
    del(gns, sgn)
    gc.collect()

    DF = DF[mask]
    del(mask)
    gc.collect()

    DF.dtype=np.dtype([('ParticleIDs', 'u8'), ('Coordinates_0', '<f8'), \
                       ('Coordinates_1', '<f8'), ('Coordinates_2', '<f8'), \
                       ('Mass', '<f8'), ('itype', 'i1')])

return DF

```

## Galaxy\_to\_past Function

```

""" # Galaxy_to_past """

def Galaxy_to_past(GalaxyID, snap_1, snap_2):
    """
    # Function to obtain center and the radio of the sphere with all
    the particles at snap_2 of the galaxy at snap_1

    snap_1 → Galaxy snapshot
    snap_2 → Past snapshot
    GalaxyID → ID of the galaxy at snap_1

    We only consider to use the darkmatter
    """

    # GroupNumber and SubGroupNumber of the Galaxy
    mask_gn_sgn = snap_1.catalogue['GalaxyID'] == GalaxyID
    gn, sgn = snap_1.catalogue[['GroupNumber',
                               'SubGroupNumber']][mask_gn_sgn][0]

    del(mask_gn_sgn)
    gc.collect()

    gns = snap_1.read_dataset(itype=1, att='GroupNumber')['GroupNumber']
    sgn = snap_1.read_dataset(itype=1, att='SubGroupNumber')['SubGroupNumber']
    mask_1 = np.logical_and(gns == gn, sgn == sgn)

    # Array with the IDs of the galaxy at snap_1
    n_particles = np.sum(mask_1)
    ParticleIDs_1 = snap_1.read_dataset(itype=1, att='ParticleIDs')['ParticleIDs'][mask_1]

    del(gns, sgn, mask_1)
    gc.collect()

    # All IDs Particles at snap_2
    ParticleIDs_2 = snap_2.read_dataset(itype=1, att='ParticleIDs')['ParticleIDs']
    # Mask to looking where are the galaxy's particles on snapshot 2.
    mask_2 = np.in1d(ar1 = ParticleIDs_2, ar2 = ParticleIDs_1)

    del(ParticleIDs_1, ParticleIDs_2)
    gc.collect()

    # Their Coordinates
    Coordinates = snap_2.read_dataset(itype=1, att='Coordinates')[mask_2][:,0]
    del(mask_2)
    gc.collect()
    Coordinates = snap_2.periodicity(Coordinates, Coordinates[0].copy())

    # Their mass
    Mass_T = snap_2.mDM*n_particles
    del(n_particles)
    gc.collect()

    # Center of mass
    Mass_center = np.array([np.sum(Coordinates['Coordinates_%i'%(i)])*snap_2.mDM for i in [0, 1, 2]])/Mass_T
    del(Mass_T)

```

```

gc.collect()

# the radios of the galaxy's particles with the new mass center
Radios = np.sum([(Coordinates['Coordinates_%i'%(i)]-Mass_center[i])**2 for i in [0, 1, 2]], axis=0)
del(Coordinates)
gc.collect()

# Maximum radio of the 90$ nearest particles.
Radio = max(np.sort(Radios)[:int(len(Radios)*0.9)])
del(Radios)
gc.collect()

return Radio, Mass_center

```

## Overrho Function

```

""" # Overrho """

def Overrho(snap, Radio, center):
    """
        Function to obtain the over-density of the sphere given as input
        (radio and center) at the snapshot also selected (snap)
    """

    with h5py.File('%s/%s.%i.hdf5'%(snap_0.path, snap_0.subpath, 0), 'r') as f:
        n_particles = f['Header'].attrs.get('NumPart_Total')
        Radios_2 = np.ones(np.sum(n_particles), dtype=np.dtype('<f8'))

    # radius of gas and DM particles with respect the center of mass.
    count = 0
    for i in [0, 1]:
        Coordinates = snap.read_dataset(itype=i, att='Coordinates')
        Coordinates = snap.periodicity(Coordinates, center)
        Radios_2[count:count+n_particles[i]] = np.sum([(Coordinates['Coordinates_%i'%(i)] \
            -center[i])**2 for i in [0, 1, 2]], axis=0)[:,0])

        del(Coordinates)
        gc.collect()

        count += n_particles[i]

    del(i)
    gc.collect()

    # Total mass, density and overdensity inside of the sphere
    Mass = np.sum(snap.particles_prop(att='Mass')['Mass'][Radios_2 <= Radio])
    rho = Mass*1.989e43/((4/3)*np.pi*((Radio)**(1/2)*3.08e24)**3)
    overrho = (rho - snap.rho_matter)/snap.rho_matter
    return overrho, rho, Mass

```

## AngularMomentum Function

```

""" # AngularMomentum """

def AngularMomentum(snap, Radio, center):
    """
    Function to obtain the total vector angular momentum of the sphere
    given as input (radio and center) at the snapshot also selected (snap)
    """
    with h5py.File('%s/%s.%i.hdf5'%(snap_0.path, snap_0.subpath, 0), 'r') as f:
        n_particles = f['Header'].attrs.get('NumPart_Total')
        Radios_2 = np.ones(np.sum(n_particles), dtype=np.dtype('<f8'))

    count = 0
    # radius of gas and DM particles with respect the center of mass.
    for i in [0, 1]:
        Coordinates = snap.read_dataset(itype= i, att='Coordinates')
        Coordinates = snap.periodicity(Coordinates, center)
        Radios_2[count:count+n_particles[i]] = np.sum([(Coordinates['Coordinates-%i'%(i)] \
            -center[i])**2 for i in [0, 1, 2]], axis=0)[:,0])

        del(Coordinates)
        gc.collect()

        count += n_particles[i]

    del(i)
    gc.collect()

    count = 0
    count_angular = 0
    angulars = np.ones((np.sum(Radios_2 <= Radio), 3), dtype=np.dtype('<f8'))
    # Calculus the vectorial product between the coordinates and the velocity of the particles inside the region
    for i in [0, 1]:
        mask = [Radios_2[count:count+n_particles[i]] <= Radio][0]
        Coordinates = snap.read_dataset(itype= i, att='Coordinates')[..., 0][mask]
        Coordinates = snap.periodicity(Coordinates, center)
        Velocity = snap.read_dataset(itype= i, att='Velocity')[..., 0][mask]

        angulars[count_angular:count_angular + np.sum(mask)][..., 0] = Coordinates['Coordinates_1']*Velocity['Velocity_2'] \
            - Coordinates['Coordinates_2']*Velocity['Velocity_1']
        angulars[count_angular:count_angular + np.sum(mask)][..., 1] = Coordinates['Coordinates_2']*Velocity['Velocity_0'] \
            - Coordinates['Coordinates_0']*Velocity['Velocity_2']
        angulars[count_angular:count_angular + np.sum(mask)][..., 2] = Coordinates['Coordinates_0']*Velocity['Velocity_1'] \
            - Coordinates['Coordinates_1']*Velocity['Velocity_0']

        del(Coordinates, Velocity)
        gc.collect()

        count += n_particles[i]
        count_angular += np.sum(mask)

    # times DM mass for DM particles
    if i == 1:
        angulars[count_angular:count_angular + np.sum(mask)] *= snap.mDM

    # times gas mass for gas particles
    elif i == 0:
        mass = snap.read_dataset(itype= i, att='Mass')['Mass'][mask][0]
        angulars[count_angular:count_angular + np.sum(mask)] *= mass
        del(mass)

    del(mask)

    del(i, Radios_2)
    gc.collect()

    # Total Angular Momentum
    angular = np.array([np.sum(angulars[i]) for i in [0, 1, 2]])
    return angular

```



## An example: How to use them

```

"""# Example storing the data

Example of how to use all the functions to obtain the dataset used in this
thesis. It consists of three parts:

The first part stores the properties of the galaxies at snapshot 28 (z= 0).

The second part stores the properties of both the formation region at
snapshot 0 (z=20) and the overdensity found in this region.

The third part stores the total angular momentum of this region. For this last
part, we also want to show how it would be done in the case where the centre of
mass and radius data of the region to be studied are already stored in the file.
"""

# Load Snapshots
snap_0 = Data_snapnum(simulation=sim, snapnum=0,
                      path='%s/snapshot_000_z020p000'%(sim))
snap_28 = Data_snapnum(simulation=sim, snapnum=28,
                       path='%s/snapshot_028_z000p000'%(sim))

n_galaxies = len(snap_28.catalogue)
print("%s: %s_disc_massive_galaxies: %s"%(sim, n_galaxies))

# Creation of the h5 file for all the information needed
h5f = h5py.File('TFG%s.h5'%(sim), 'a')
# Creation/use of the simulation part of the .h5py dataset
try:
    Idf = h5f.create_dataset(sim, data=snap_28.catalogue['GalaxyID'])
except:
    Idf = h5f['%s'%(sim)]

# First Part
Idf.attrs['Mass'] = snap_28.catalogue['Mass']
Idf.attrs['MassStar'] = snap_28.catalogue['MassType-Star']
Idf.attrs['Vmax'] = snap_28.catalogue['Vmax']
Idf.attrs['GasSpin_z'] = snap_28.catalogue['GasSpin_z']
Idf.attrs['KappaCoRot'] = snap_28.catalogue['KappaCoRot']

print("Save_1st_Info_-->_Done")

# Second part
radio = np.zeros(n_galaxies, dtype= np.dtype('<f8'))
mass_center = np.zeros((n_galaxies, 3), dtype= np.dtype('<f8'))
Mass = np.zeros(n_galaxies, dtype= np.dtype('<f8'))
rho = np.zeros(n_galaxies, dtype= np.dtype('<f8'))
overrho = np.zeros(n_galaxies, dtype= np.dtype('<f8'))

for i in range(n_galaxies):
    radio[i], mass_center[i] = Galaxy_to_past(GalaxyID = snap_28.catalogue['GalaxyID'][i],
                                             snap_1 = snap_28, snap_2 = snap_0)
    overrho[i], rho[i], Mass[i] = Overrho(snap = snap_0,
                                          Radio = radio[i],
                                          center = mass_center[i])

Idf.attrs['radio_sphere'] = radio
Idf.attrs['center_mass'] = mass_center
Idf.attrs['Snap_0_mass'] = Mass
Idf.attrs['Snap_0_rho'] = rho
Idf.attrs['Snap_0_overrho'] = overrho

print("Save_2nd_Info_-->_Done\n")

# Third part
atrib = lambda att: Idf.attrs.get(att)
angular = np.zeros((n_galaxies,3), dtype= np.dtype('<f8'))

for i in range(n_galaxies):
    angular[i] = AngularMoment(snap = snap_0,
                              Radio = atrib('radio_sphere')[i],
                              center = atrib('center_mass')[i])

Idf.attrs['Snap_0_angular'] = angular
print("\nSave_3rd_Info_-->_Done")

h5f.close()

print(input('\n_Exist:'))

```