CTA200H Mini-Project: MW Analogues in Cosmological Simulations

Alicia Savelli

Advisors: Ted Mackereth and Josh Speagle

May 15, 2021

1 Introduction

The star-forming Galaxy Main Sequence (GMS) is a current and fairly recent topic of study in astronomy. The GMS comes out of the relationship between galactic Stellar Mass (SM) and Star Formation Rate (SFR), as illustrated in Figure 1. In this model, the GMS is represented by the blue-shaded central grouping of galaxies along which majority of star-forming galaxies are thought to lie. The remaining galaxies fall into the outlier groups 'starburts', 'the green valley', and 'red and dead'. The galaxies in the starbursts grouping have an exceptionally fast rate of star-formation; they are plotted in pink. The low-rate star-forming galaxies are found in the green valley and the read and dead galaxies, pictured in green and red, respectively. One current understanding of
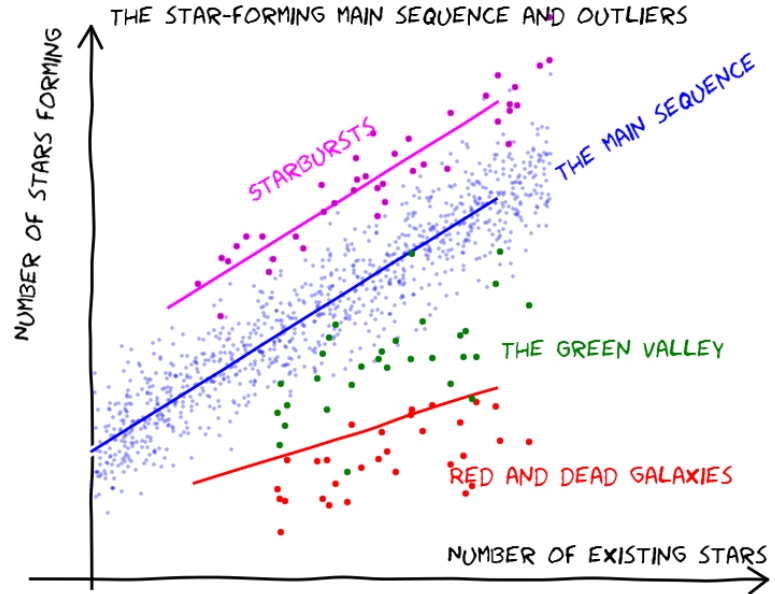

Figure 1: Cartoon sketch of the GMS and outliers. Majority of galaxies are pictured along the blue main sequence. The outliers are grouped into high-rate star-forming galaxies ('starbursts') laying above the GMS in pink and low-rate star-forming galaxies ('red and dead') falling below the GMS in red [1].

1

the GMS is that as the SM of a galaxy grows, so too does its SFR, until the dust needed for star-formation has been used up and the galaxy is said to be 'quenched'. At this stage, the galaxy 'falls off' the main sequence and enters the green valley and the red and dead group, where star-formation is very low [1].

The name 'GMS' is reminiscent of the well-known stellar main sequence, although its relationship is not as well established. The above interpretation is only one model of the GMS, and other studies have suggested both that the main sequence and the red and dead galaxies may be two separate groupings of galaxies, and that the groups may lie along a single smooth distribution [2]. Another factor to consider in the study of the GMS is the morphology of the galaxies along the distribution in relation to the Hubble sequence [2].

The purpose of this mini-project was to become familiar with writing `SQL` queries to pull data from the EAGLE galaxy formation simulations and analyze their output using `matplotlib` in `python`. First, the SM and SFR of galaxies at redshift $z = 0$ were plotted against each other in a variety of styles. The graphs were analyzed and their features compared to physical observations of the GMS. Then, SM and SFR of galaxies at redshift $0 < z < 0.5$ were plotted and analyzed. This range in redshift corresponds to a lookback time of 5.19 billion years, and thus the data illustrates the formation and evolution of these galaxies through the merging of smaller galaxies at longer redshifts.

2   Methods and Results

The study of the GMS was completed in two parts: analysis of data from the EAGLE simulations at $z = 0$ and at $0 \leq z < 0.5$. A description of the simulation and the data in the tables can be found in Reference [3]. The script used to generate the figures and analyze the data can be found at `https://github.com/as14bq/CTA200/tree/master/project`.

2.1   Retrieving EAGLE Galaxy properties: your first `SQL` query in `python`

The initial query was coded as:

```
SELECT sub.MassType_Star as stellar_mass,
       sub.StarFormationRate as SFR,
       mag.r_nodust as r,
       mag.g_nodust as g
FROM RefL0100N1504_SubHalo as sub,
     RefL0100N1504_Magnitudes as mag
WHERE sub.SnapNum = 28 and
      sub.GalaxyID = mag.GalaxyID
```

which imported the columns for SM, SFR, and the rest-frame absolute magnitudes in the g and r bands without dust attenuation from the `RefL0100N1504_SubHalo` and `RefL0100N1504_Magnitudes` tables for a `Redshift` $z = 0$, and joined the tables on `GalaxyID`.
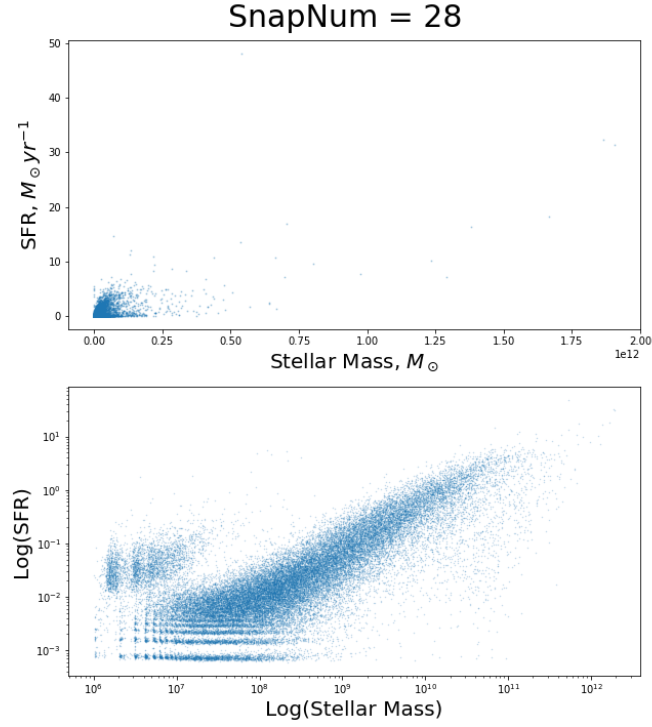
Figure 2: Plot of the galactic Stellar Mass (SM) ($M_\odot$) against Star Formation Rate (SFR) ($M_\odot \mathrm{yr}^{-1}$) on linear scales (above) and logarithmic scales (below) at `SnapNum = 28` ($z = 0$). A relationship between SM and SFR becomes visible in the loglog plot of the data. The points have been made partially transparent so that the density of galaxies across the plot can be more easily observed.
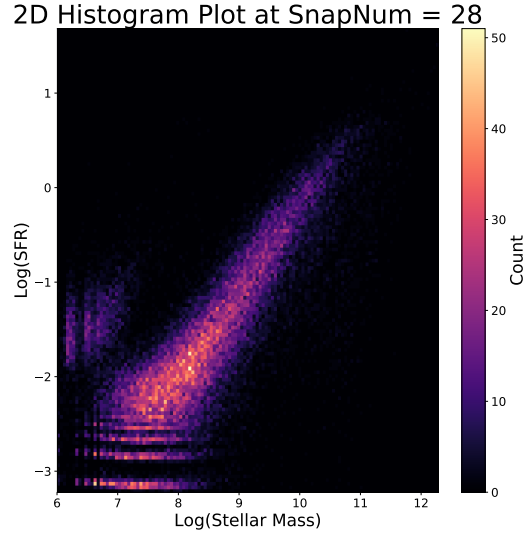
Figure 3: Loglog histogram of the galactic Stellar Mass (SM) ($M_\odot$) against Star Formation Rate (SFR) ($M_\odot \mathrm{yr}^{-1}$) at `SnapNum = 28` ($z = 0$). The figure shows a higher density of galaxies along the centre of the main sequence.
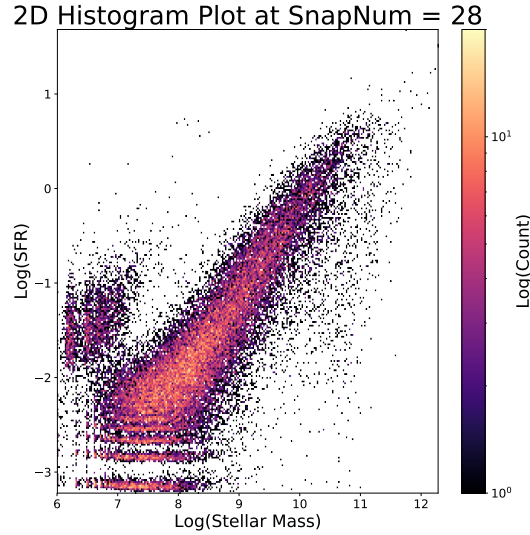

Figure 4: Loglog histogram of the galactic Stellar Mass (SM) ($M_\odot$) against Star Formation Rate (SFR) ($M_\odot \mathrm{yr}^{-1}$) with a logarithmic count at `SnapNum = 28` ($z = 0$). The figure shows a higher density of galaxies along the centre of the main sequence.
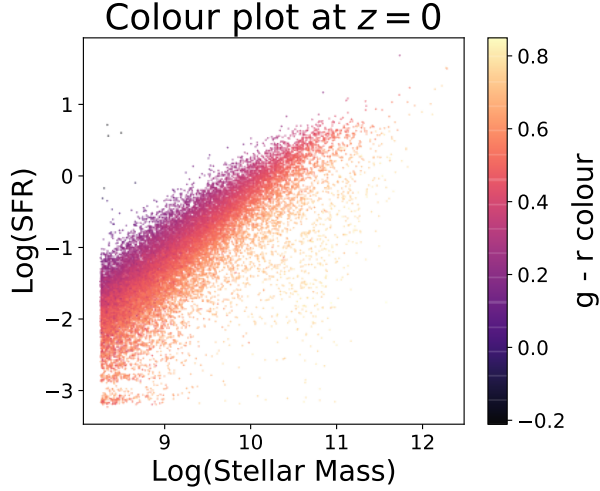
Figure 5: Loglog plot of the galactic Stellar Mass (SM) ($M_\odot$) against Star Formation Rate (SFR) ($M_\odot\mathrm{yr}^{-1}$) coloured by $g - r$ colour at $z = 0$. The figure shows redder galaxies towards the bottom of the sequence and greener galaxies above.

The script plotted the galactic Stellar Mass (SM) ($M_\odot$) against Star Formation Rate (SFR) ($M_\odot\mathrm{yr}^{-1}$) on both linear and logarithmic scales (Figure 2). The relationship between SM and SFR is not immediately obvious on the linear scaling, but with logarithmic axes there appears to be a sequence of galaxies where SFR increases with SM. Although the opacity of the points was adjusted to give a better representation of the density, more representative loglog histograms of the data are presented in Figures 3 and 4, where the density is indicated by the colouring on a linear scale and a logarithmic scale, respectively. These histogram plots give a better indication of the number of galaxies along the GMS. A number of key features can be observed in all three of the loglog plots, as well as a few differences between them, all of which will be expanded upon in the Discussion section.

A final analysis of the data taken at $z = 0$ was a plot of the colour of the galaxies along the GMS. The difference between the magnitudes of the galaxies in the green and red bands was taken to provide the colouring of the loglog plot in Figure 5.

2.2 Going Deeper into the Simulations: more properties and more snapshots

The second part of the study examined galaxies at `Redshift` $0 \le z < 0.5$. The query was coded as:

```
SELECT MassType_Star as stellar_mass2,
       StarFormationRate as SFR2,
       GalaxyID as galaxyID,
       TopLeafID as topleaf,
```

```
        LastProgID as lastprog,
        DescendantID as descID,
        Redshift as Z,
        Mass as mass
  FROM RefL0100N1504_SubHalo
  WHERE Redshift < 0.5
  ORDER BY GalaxyID
```

where SM and SFR have again been pulled from table **RefL0100N1504_SubHalo**, along with the columns needed for analyzing the progenitors, the `Redshift`, and the overall mass, for $0 \leq z < 0.5$, and the results were ordered by their `GalaxyID`.

The same loglog plots from the first part of the study were run again for this higher volume of galaxies, with results illustrated in Figures 7, 8, and 9. Similarities and differences between these same plots at $z = 0$ and at $0 \leq z < 0.5$ will be explored in the Discussion section.

With a larger range in redshift, the same galaxies can be analyzed across different times. Figure 6, taken from [3], shows an example of a merger tree at $z = 0.18$. A merger tree shows all the past history of a given galaxy. All of the past galaxies that have merged to give the present galaxy (progenitor galaxies) can be found from querying all galaxies with an ID between the given `GalaxyID` and its `LastProgID`. The galaxy chosen to be analyzed was the most massive galaxy, of SM 1.9054253×10^{12} M$_\odot$ and ID 21379521. Analysis of the data found that there were 27 descendants of Galaxy 21379521 along the main branch, and 194064 total descendants.

3   Discussion and Conclusions

`SQL` queries work my retrieving columns of data from specified tables under given conditions. In the first query used in this study, the information retrieved from the simulation was the SM and SFR, under the condition `SnapNum = 28` (corresponding to $z = 0$). Alternatively, this condition could have been coded as `WHERE Redshift = 0`. Importing data from one `SnapNum` means that the data output will be from one snapshot in time.

The loglog plots in Figures 2, 3, and 4 share a number of features that are both supported by and distinct from observational studies. A clear relationship can be seen between the SM and SFR in the form of a thick central sequence going from the lower left of the plots up towards the upper right. This represents the GMS. Additionally, a cluster of galaxies can be viewed above the GMS, which agrees with observational studies of starburst galaxies. The unusual features shown in these plots are the vertical and horizontal bands seen in the bottom left of the plots. The reason for these bands is a result of a simplification used to make it possible to run such large volume simulations of the Universe. The simulation makes use of subgrid models, which assume a simple model to reproduce the processes that are too small to be properly simulated, but still have long-lasting effects on the formation of galaxies. Subgrid models are described by parameters constrained directly from obser-
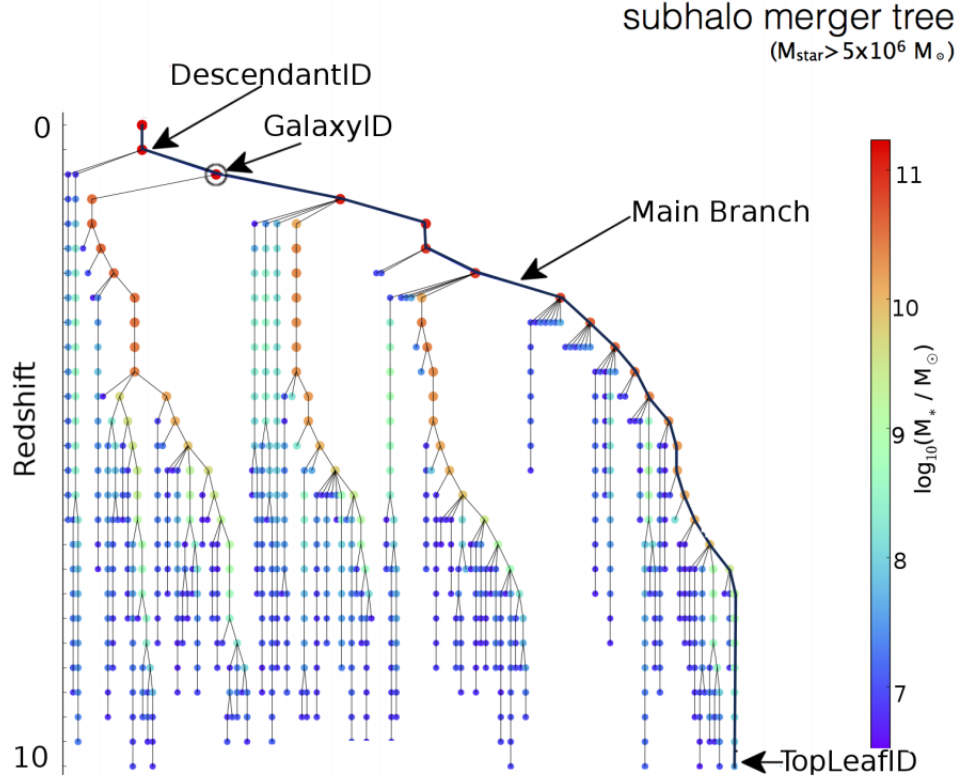
Figure 6: Merger history of a galaxy with a $z = 0.18$ stellar mass $M_* \sim 10^{10}$ $M_\odot$ indicated by the circled dot. Symbol colours and sizes are scaled with the logarithm of the stellar mass. The `GalaxyID` of this galaxy points towards it, as indicated by the arrow. The main progenitor branch is indicated with a thick black line, all other branches with a thin line. The `TopLeafID` gives the `GalaxyID` of the highest redshift galaxy on the main progenitor branch whilst the `LastProgID` (not shown) gives the maximum `GalaxyID` of all the progenitors of the galaxy considered. Querying all galaxies with an ID between `GalaxyID` and `LastProgID` will return all the progenitor galaxies in the tree. (Taken verbatim from [3])
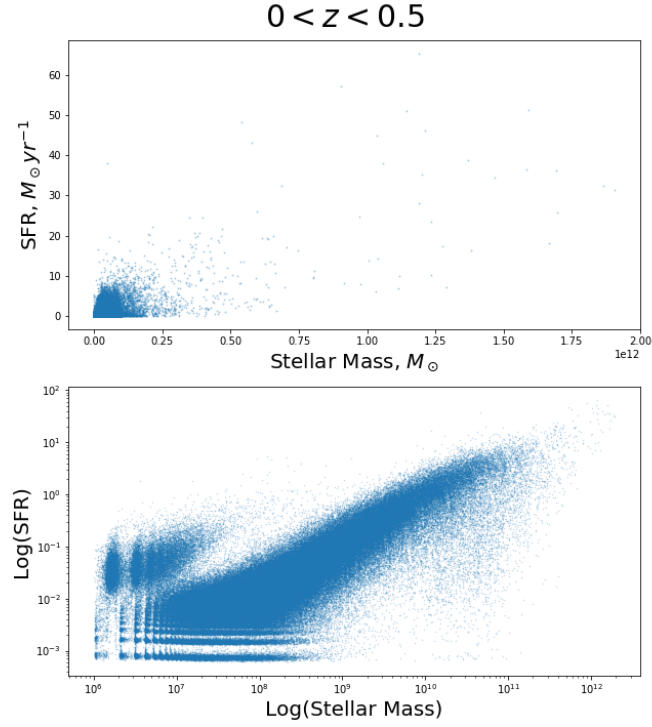
Figure 7: Plot of the galactic Stellar Mass (SM) ($M_\odot$) against Star Formation Rate (SFR) ($M_\odot \text{yr}^{-1}$) on linear scales (above) and logarithmic scales (below) at $0 \leq z < 0.5$. A relationship between SM and SFR becomes visible in the loglog plot of the data. The points have been made partially transparent so that the density of galaxies across the plot can be more easily observed.
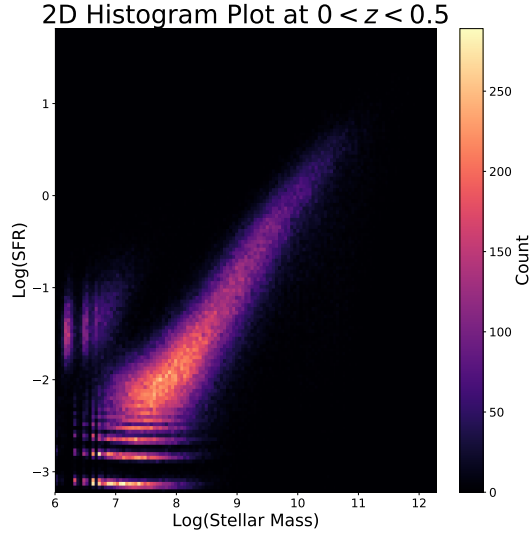
Figure 8: Loglog histogram of the galactic Stellar Mass (SM) ($M_\odot$) against Star Formation Rate (SFR) ($M_\odot \text{yr}^{-1}$) at $0 \leq z < 0.5$. The figure shows a higher density of galaxies along the centre of the main sequence.
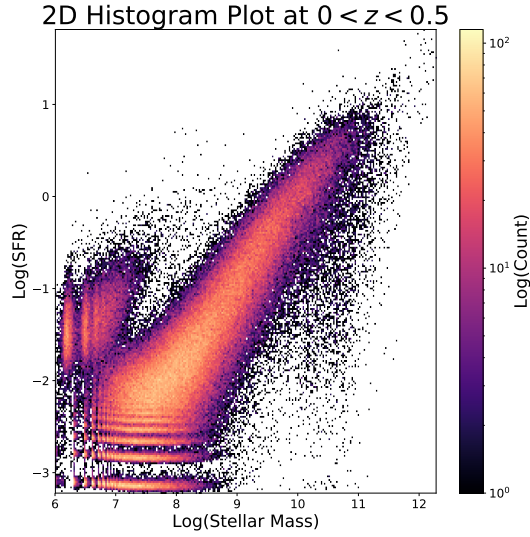

Figure 9: Loglog histogram of the galactic Stellar Mass (SM) ($M_\odot$) against Star Formation Rate (SFR) ($M_\odot \text{yr}^{-1}$) with a logarithmic count at $0 \leq z < 0.5$. The figure shows a higher density of galaxies along the centre of the main sequence.

9

vations or calibration data [4]. The bands in these plots suggest that there must not be effective calibration data or known parameters for lower mass galaxies.

When considering galaxies from $0 \leq z < 0.5$, there is a significant increase in galaxies being analyzed. The first query pulled 2,275,510 galaxies (41,491 after taking the logarithm), and the second pulled 11,819,412 galaxies (269,083 after taking the logarithm). The higher volume of galaxies made the plots much denser along the main sequence and also expanded the area of the graph covered by the main sequence.

Selecting galaxies across a range of redshifts is more representative of what would be seen in an observational galaxy survey possibly because in reality, the galaxies would be evolving at different stages and would be at different distances and redshifts when they are observed rather than a single clean snapshot. The maximum redshift that could be observed would be influenced by a number of factors such as if the galaxy is too far for the light to reach our telescopes (i.e., past the observable universe), if redshift causes the light to shift so far along the EM spectrum that our instruments are not capable of picking up the signal, and the amount of time the survey is left to run for.

References

[1] P. Alton, *The end of the line*, Accessed: 2021-05-11, Astrobites, 2016, `https://astrobites.org/2016/11/10/the-end-of-the-line/`.

[2] S. Eales, P. de Vis, M. W. L. Smith, K. Appah, L. Ciesla, C. Duffield, and S. Schofield, "The galaxy end sequence", Monthly Notices of the Royal Astronomical Society **465**, 3125–3133 (2016), `http://dx.doi.org/10.1093/mnras/stw2875`.

[3] S. McAlpine, J. Helly, M. Schaller, J. Trayford, Y. Qu, M. Furlong, R. Bower, R. Crain, J. Schaye, T. Theuns, and et al., "The eagle simulations of galaxy formation: public release of halo and galaxy catalogues", Astronomy and Computing **15**, 72–89 (2016), `http://dx.doi.org/10.1016/j.ascom.2016.02.004`.

[4] *The EAGLE project*, Accessed: 2021-05-11, Institute for Computational Cosmology, 2016, `http://icc.dur.ac.uk/Eagle/about.php`.