On Disk Galaxies in IllustrisTNG: Insights on the Main Sequence, Mass-Size and Tully-Fisher Relations

Abstract

This study explores a method of discerning disk galaxies from the IllustrisTNG Simulation. Our method involves filtering galaxies based on their alignment with two empirical relations: the Star-Forming Main Sequence(SFMS) and the Mass-Size Relation with regards to the permitted scatter. Subsequently we revisit the redshift evolution of these relations and attempt to explain our observations from a physical standpoint wherever possible.

To validate the efficacy of this method we perform a simple comparison of the results with the Baryonic Tully Fisher Relation and find a satisfactory level of agreement. We also generate a random subset of images from these mock catalogs and find that they are disks.

We then explore the limitations inherent to our methodology and propose modifications that could be implemented so as to improve the accuracy of selection.

Lastly we release our catalogs publicly for use.

Contents

1 Introduction

1.1 Galaxy Morphology

The number of galaxies in the observable universe exceeds 200 billion which exhibit a multitude of properties. Given this vast count, it is also desirable to look for these patterns from an organisational and statistical standpoint and perform systematic analyses to do meaningful science. In this study we focus on one aspect of these trends, i.e., morphology.
Morphological classification of galaxies is done based on their structure and visual appearance. One way to do this is to consider features such as the presence of spiral arms, the central bulge, and the overall structure of the galaxy. This is well incorporated in the century-old Hubble Classification Scheme.

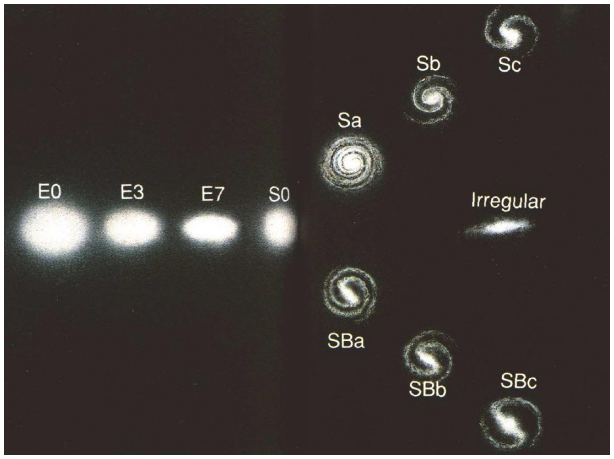
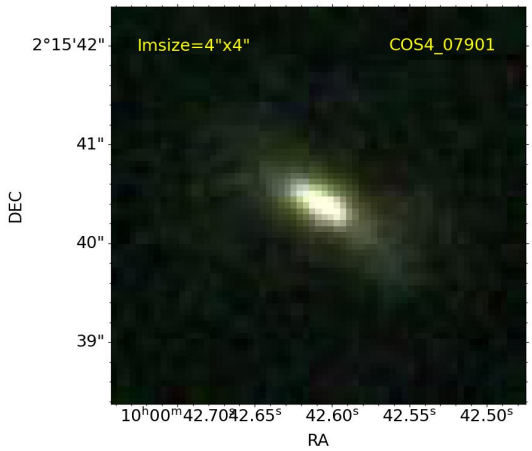


Figure 1: The morphological sequence of galaxies predicted by Hubble. E0,E3 and E7 are different classes of elliptical galaxies, referred to as early-type galaxies. Sa,Sb,Sc,SBa,SBb and SBc are spiral galaxies, also known as late-type galaxies.*Early and late do not imply evolutionary sequence.*

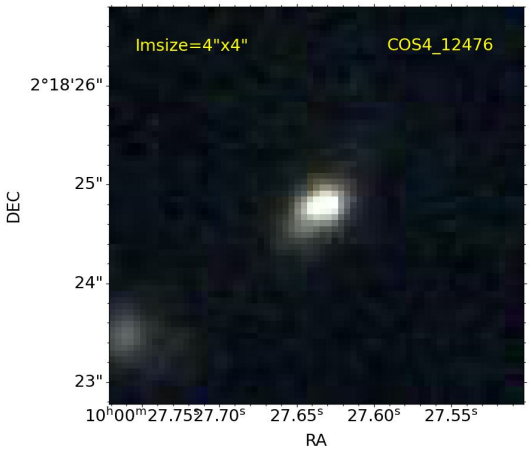
Known colloquially as the 'Hubble Tuning Fork,' the Hubble sequence divides galaxies into three broad classes:

1. The Ellipticals, featureless galaxies appearing as ellipses,
2. The Spirals, consisting of a bulge and a flattened disk with spiral arms(and sometimes, bars),
3. The lenticulars consisting of a distinct disk and bulge but no spiral arms.

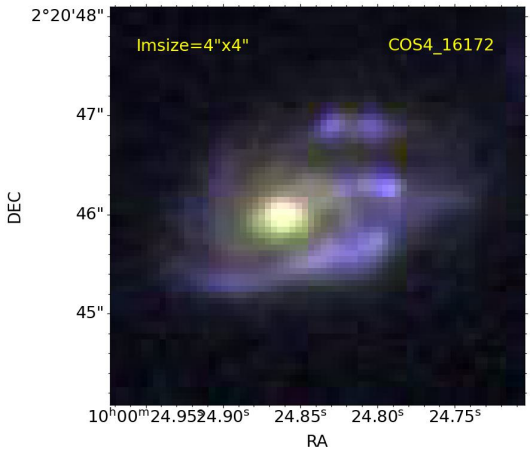
Situated outside the tuning fork are irregular galaxies. No particular shape can be attributed to them.



(a) An elliptical galaxy; A flat galaxy, without any bulge.



(b) A lenticular galaxy; The disk is visible with a hint of a bulge.



(c) A Barred Spiral Galaxy; The bars emit lots of blue light indicating active star formation.

Figure 2: RGB Image Cutouts of 3 Galaxies from the HST COSMOS FITS Files. The disk-like feature is visible, faintly though, in (c).

The morphology of a galaxy is a consequence of its formation and evolution as dictated by stellar population, star formation rate, internal perturbations including supernova and Active Galactic Nuclei(AGN) feedback, merger histories, and the influence of the dark matter halo they reside in.
Studying galaxy morphology is not just a window into their evolution history; it conveys information about dark matter distribution [1] and provides tests for modified gravity theories [2] among other implications for cosmology. Thus, although morphology is an age-old facet of galactic science, it is appropriate and relevant to present-day research.

1.2 Objective

The objective of this study is to extract disk galaxies from the IllustrisTNG Simulation, the most extensive high-resolution simulation of the galaxy formation to date. The galaxies will be stored as a comprehensive catalog of their significant physical properties, such as SFRs, sizes, component masses, and velocity dispersion, which can then be studied for agreement with empirical observations, trends, and so on.

1.3 Methodology

Modern classification approaches revolve around machine learning algorithms incorporating deep learning for feature extraction and image recognition in a dataset. However, it has also been proved over years of research that the human eye has a better aptitude for this, so crowdsourcing and citizen science via platforms like the Galaxy Zoo Project [3] is also a popular methodology in modern science.

In our research, however, we resort to observational properties of disk galaxies to extract them. This method can be more robust and valuable for this project because, in addition to the classification itself, we can also compare the simulation results with observations. It will not only give more credibility to our results but also bring out potential disparities between observation and the simulation.

The observational scaling relations that form the basis of our methodology are the *Star-Forming Main Sequence* and the *Mass-Size Relation*.

1.4 Star-Forming Main Sequence

With reference to their star formation activities, galaxies generally fall into two categories: star-forming and quiescent. Star-forming galaxies largely include disks. Being younger, they show active star formation in their spiral arms(in case of spiral galaxies) or their disks(S0 Galaxies).

At a fixed redshift, the Star Formation Rates(SFRs) of these star-forming galaxies and their stellar masses are known to exhibit a tight correlation called the star forming main sequence. Galaxies that lie above the main sequence relation are called starburst galaxies: star formation rates as high as these are usually triggered by merger events. Galaxies that lie well below this relation are called quiescent galaxies, they have likely exhausted their cold gas for star formation or have low SFRs due to AGN feedback. They are mostly ellipticals [4]. This can be explained considering that current evidence points at elliptical galaxies being formed from spiral galaxy mergers, and are thus dominated by much older stars.

There are two divergent viewpoints on the exact form of this relation. Some studies

report a linear relation between $\log M_*$ and $\log SFR$ (Whitaker et al 2012[5],Speagle et al 2014[6], Iyer et al 2018[7]), more recent studies also report a turnover in this relationship at the high mass end, at $M \approx 10^{10} M_\odot$(Lee et al 2015[8],Tomczak et al 2016[9] and Leja et al 2022[10]). There is thus no consensus on the exact shape of the main sequence.

Schreiber et al 2015[11] asserts that the reason could be the contribution of the bulge to the stellar mass whereas majority of the SFR comes from the disk. Johnston et al 2015[12] suggests that this disparity could be the result of selection; selection based on the D4000 index yields a power law relation between the two and a color-color selection yields a distinct turnover at the high-mass end, further suggesting that this is a result of contamination by passive galaxies(see also [13]). Our study attempts to extract star-forming disks, thus prompting us to stick to the power law formulation:

$$\log SFR = \alpha \log M_* + \beta$$

α is the power law index, or the slope of the main sequence and β the normalisation. The relationship exists for about 4 orders of magnitude in stellar mass [14] up until $z = 6$ [6]. The observed scatter is reported by multiple studies to be around 0.3 dex, however the results are conflicting in that scatter is in some cases reported to be constant across stellar masses and redshifts (Speagle et al 2014[6], Schreiber et al 2015[11]); some studies report an increase in scatter towards higher masses (Santini et al 2017[15]). The origin of intrinsic scatter is believed to be repeated cycles of flow of material into galaxies leading to wet compaction followed by inside out quenching, until departure from the MS once the disc replenishment time is longer than the core depletion time, thus achieving quiescence [16]. The total observed scatter takes into account other systematic uncertainties associated with observations. The observational results we use in our study all exhibit constant scatter throughout the mass range under their consideration, unless otherwise stated.

The observation that younger stars are bluer might misleadingly suggest that UV is a better SFR indicator, however UV emission also interacts with the surrounding dust by heating it up leading to re-emission in the IR regime, thus underestimating the SFR (Santini et al 2009[14], Whitaker et al 2012[5]) .This problem can be circumvented by harnessing a combination of UV and IR measurements.

There have been multiple studies aimed at determining α and β employing several multiwavelength data with varying degrees of accuracy.
One of these works Speagle et al 2014 [6] gives a relatively reliable result accounting for differences in calibration and systematic uncertainties. Until $z = 5$, the relation is stated to be:

$$\log SFR(M_*, t) = (0.84 \pm 0.02 - 0.026 \pm 0.003 \times t) \log M_* - (6.51 \pm 0.24 - 0.11 \pm 0.03 \times t)$$

where t is the age of the universe in Gyr.

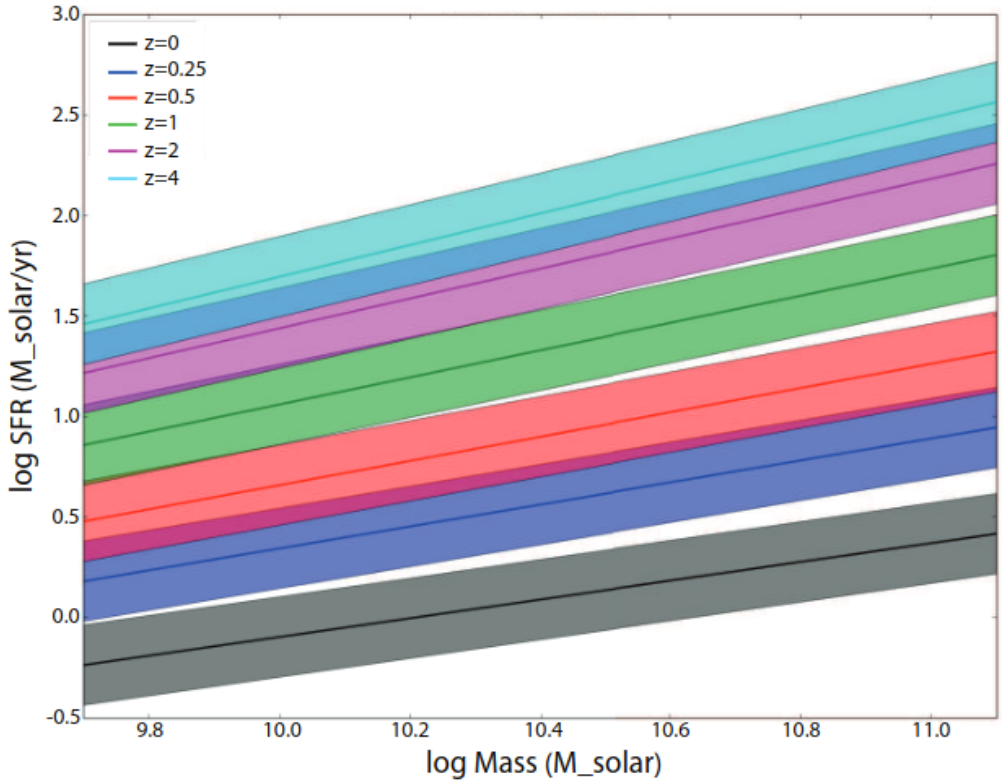


Figure 3: Main Sequence relations as in Speagle *et al* 2014 at varying redshifts. The shaded region represents the intrinsic scatter of 0.2 dex found to be constant over cosmic time.

Filtering galaxies that fall within the allowed scatter of the main sequence as a zeroth order estimate can be considered to be disks. If fits to the data are not observed to be within the permitted range of this relation, we compare with newer studies that employ better methodology/selection criteria or studies that confine themselves to the desired redshift range. These studies will naturally match the data better.The galaxies retained after this selection are then passed on to the next step for selection based on the size-mass relation.

1.5 Size-Mass Relation

Observations suggest the existence of a well-defined empirical size-mass relation for the desired redshift range. Late-types and early-types have their separate size-mass relations. For extracting galaxies based on this we use results from two publications, one for the local galaxies at $z \approx 0$, Lapi *et al* 18 [17] and another for the redshift range $0 < z < 3$, van der Wel *et al* 14[18].

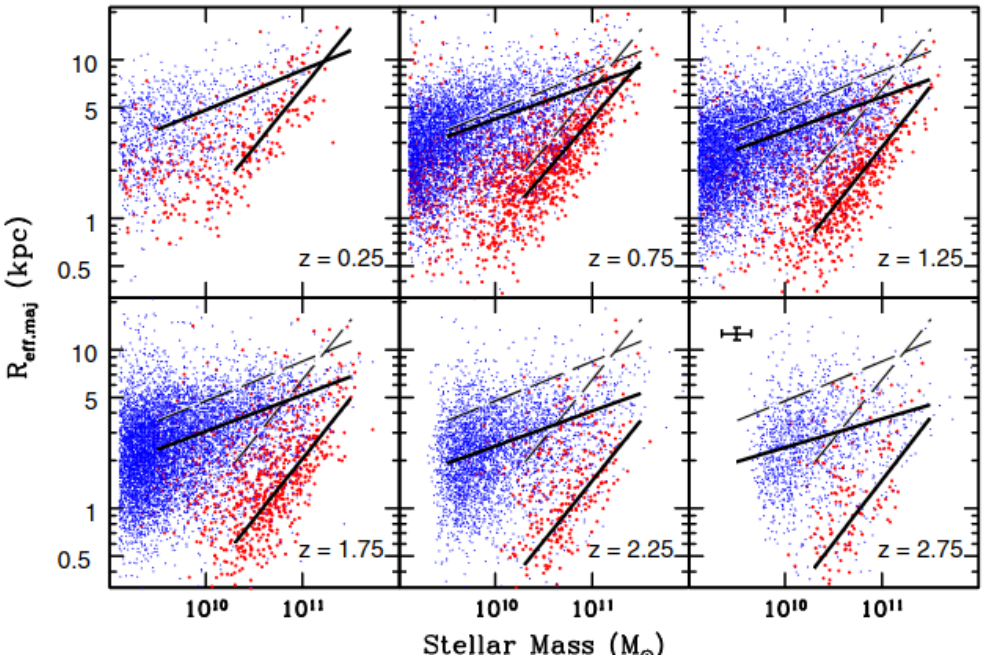


Figure 4: Size–stellar mass distribution of late(blue) and early-type (red) galaxies as in van der Wel+14

In the latter, the size mass relation is fit to a power law of the following form:

$$\frac{R_{eff}}{kpc} = A\left(\frac{M_*}{5\times10^{10}M_\odot}\right)^{\alpha}$$

. Here R_{eff} is the effective radius of the galaxy, defined to be the three-dimensional stellar half-mass radius [19], which is a good proxy for the size. The fits to various redshifts are given below.

	Early-type Galaxies			Late-type Galaxies		
z	$\log(A)$	α	$\sigma\log(R_{\rm eff})$	log A	α	$\sigma\log(R_{\rm eff})$
0.25	0.60 ± 0.02	0.75 ± 0.06	0.10 ± 0.02	0.86 ± 0.02	0.25 ± 0.02	0.16 ± 0.01
0.75	0.42 ± 0.01	0.71 ± 0.03	0.11 ± 0.01	0.78 ± 0.01	0.22 ± 0.01	0.16 ± 0.01
1.25	0.22 ± 0.01	0.76 ± 0.04	0.12 ± 0.01	0.70 ± 0.01	0.22 ± 0.01	0.17 ± 0.01
1.75	0.09 ± 0.01	0.76 ± 0.04	0.14 ± 0.01	0.65 ± 0.01	0.23 ± 0.01	0.18 ± 0.01
2.25	-0.05 ± 0.02	0.76 ± 0.04	0.14 ± 0.02	0.55 ± 0.01	0.22 ± 0.01	0.19 ± 0.01
2.75	-0.06 ± 0.03	0.79 ± 0.07	0.14 ± 0.03	0.51 ± 0.01	0.18 ± 0.02	0.19 ± 0.01

Note. $\sigma(\log R_{\rm eff})$ is the scatter in $R_{\rm eff}$ in logarithmic units.

Figure 5: Results from the Parameterized Fits to the Size–Mass Distribution as in van der Wel+14

2 Data

IllustrisTNG is a set of cosmological hydrodynamical simulations for galaxy formation including treatments of gas cooling, star formation and feedback from stellar winds, supernovae and AGN [19]. TNG50 is its latest and highest-resolution simulation. It realises a 50 Mpc box sampled by 2160^3 gas cells, dark matter and Monte-Carlo Tracer Particles each, allowing us to analyse galaxy evolution since $z = 6$ in unparalleled detail. We choose TNG50 to obtain the highest possible resolution do discern the disk structure.

There TNG Subhalo Catalog holds information on masses of gas, dark matter and stars, star metallicities, photometrics, velocity dispersions etc in galaxies. In our analysis we use fields containing stellar masses, star formation rates and stellar half-mass radii(converted from comoving to physical units) .

2.1 Selection Criteria

We reject galaxies with unreliable data on the above fields:

- `SubhaloFlag=1` Reject galaxies that may have formed within an existing halo, or are possibly baryonic fragments of a disk or other galactic structure identified so as to include only those of cosmological origin, thus picking only central galaxies.
- $0 <$`['SubhaloMassType'][:,4]` $< \infty$
- $0 <$`['SubhaloSFR'][:,4]` $< \infty$
- $0 <$`['SubhaloHalfMassRadType'][:,4]` $< \infty$

Note that 4 denotes the particle type, in this case, the stars; we thus reject unrealistic values of stellar mass, stellar half mass radius and SFR.
Further, we restrict our analysis that satisfy $8 < \log M_* < 11.5$ and $0 \leq z \leq 3.05$. In accordance with the redshifts for which fits are available in van der Wel et al 2014, we divide our redshift range into the following bins, such that the slopes and normalizations for the mass size relation at median redshifts are available to us.

Redshift Bin	**Median**
$[0, 0.05)$	N/A
$[0.05, 0.45)$	0.25
$[0.45 - 1.05)$	0.75
$[1.05 - 1.45)$	1.25
$[1.45 - 2.05)$	1.75
$[2.05 - 2.45)$	2.25
$[2.45 - 3.05)$	2.75

3 Analysis

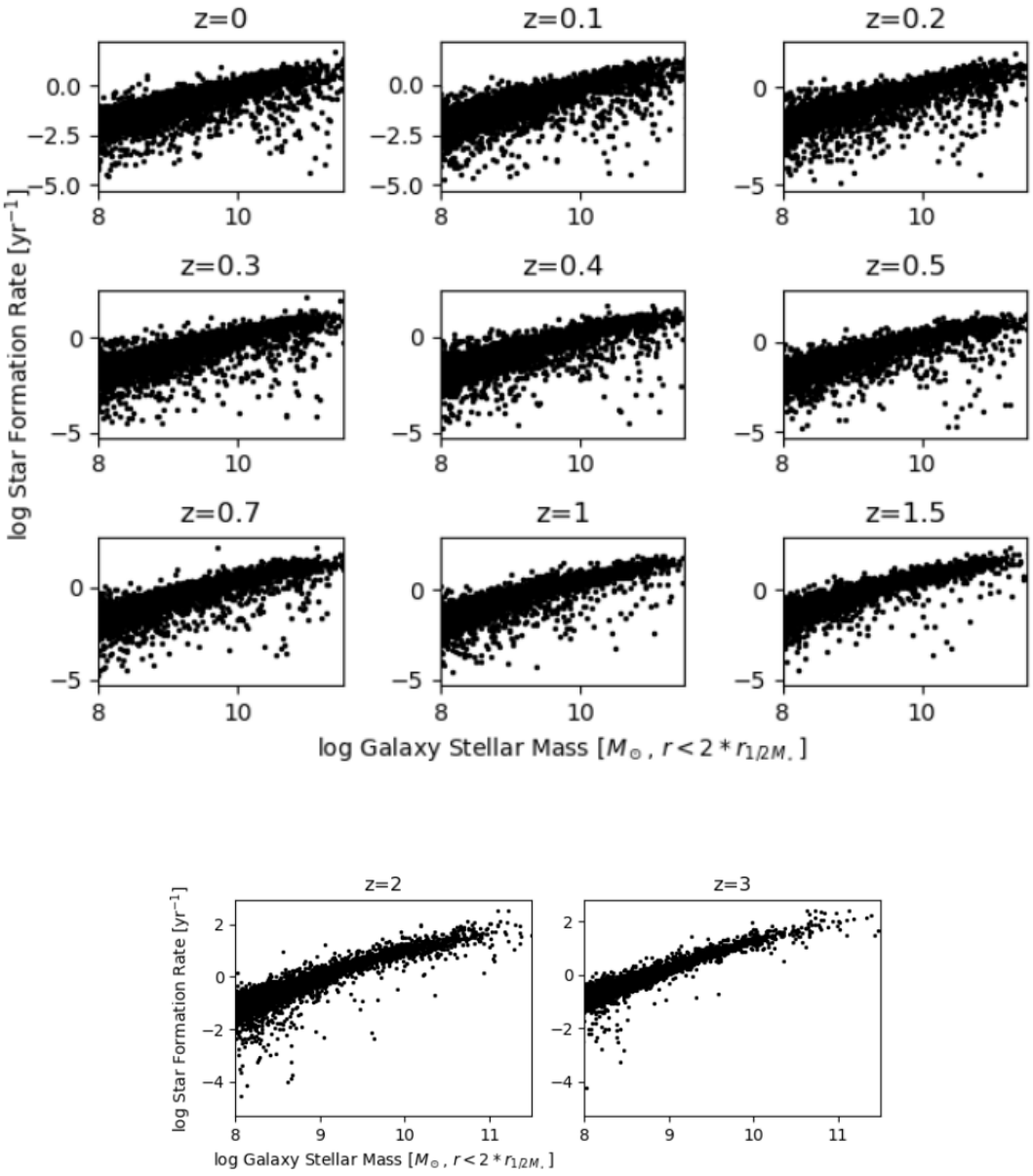


Figure 6: The distribution of $\log SFR$ against $\log M_*$ with increasing redshift. There appears to be a straight line around which maximum galaxies exist(the main sequence), alongwith a large number of outliers. These are expected to be mostly quiescent galaxies.

As a basic analysis, the general algorithm we follow for selecting galaxies according to the main sequence is the following:

1. Noting that simulations are immune to systematics whereas observational results

are not, we remove galaxies beyond (nearly)twice the total scatter(2σ) of the assumed relation to include 95% of disks.

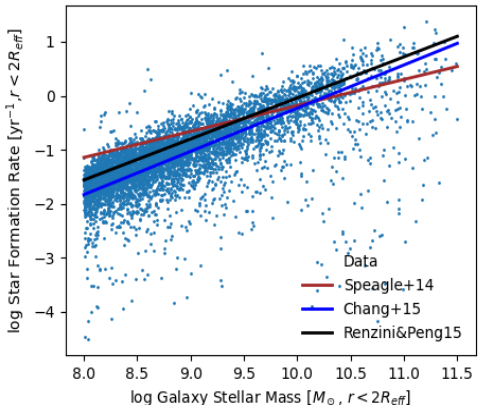
2. We bin the data and find the best fit to this binned data. If it lies within the allowed scatter, the retained galaxies are likely disks.

3. It may be possible that the best fit to the data doesn't agree with the observational relation. Essentially the slopes and normalisation are strongly dependent on the selected sample and the SFR tracer [6]. To overcome this, we take two approaches: use Speagle's Relation for all z-bins, or use local-z relations for each z-bin.

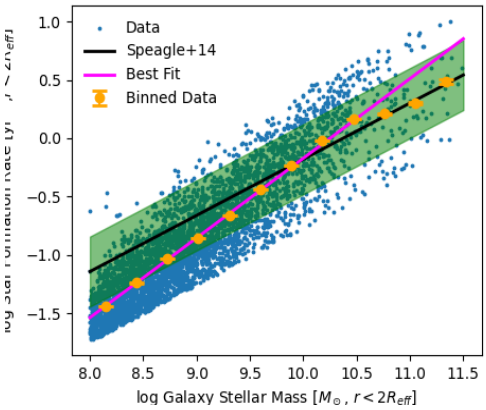
The number of galaxies itself seems to decrease drastically as one moves to higher redshifts. In order to improve the thoroughness and accuracy of the filtering we carry out the analysis for each redshift bin separately.

We show the process for one redshift bin, $0 < z < 0.05$, the other bins being a repetition of the same.

Fig 5(a) shows the comparison of three studies at z=0(one snapshot in this bin). Chang et al 2015[20](hereafter C15) covers galaxies with $z < 0.1$ and Renzini & Peng 2015[13](hereafter RP15) cover galaxies with $z < 0.085$. Both of these appear to be in better agreement with simulation data as compared to Speagle et al, since they cover a narrow redshift range. C15 provides a robust methodology of estimating the mainsequence by combining SDSS and WISE photometry and including effects of dust attenuation and emission. RP15 provides an objective definition of the main sequence immune to selection bias. Further, C15 covers the redshift range upto 0.2, whereas only 0.085 in case of RP15. In this case therefore, RP15 proves to be a more robust result. As mentioned earlier, for approach 1 we continue with Speagle's Relation and for 2 we use RP15. The same process is repeated for all snapshots within the bin.

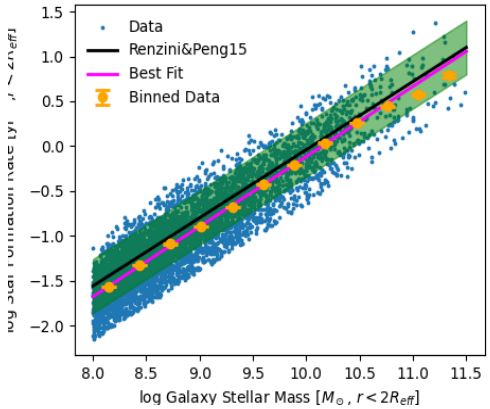


(a) Simulation data and its comparison with three studies



(b) 0.6 dex clipped and mean-binned data and corresponding best fit and standard error. 0.3 dex is the total scatter.

Figure 7: Selecting galaxies according to S14



(a) 0.6 dex clipped and mean-binned data and corresponding best fit with standard error. 0.3 dex is the total scatter.

Figure 8: Selecting galaxies according to RP15

We now combine the selected galaxies for filtering according to the size-mass relation.

Lapi et al 2018 [17] provides precision scaling relations for local disk galaxies. For the size-mass relation we digitise Figure 5 of the paper that provides a highly precise relation with a total scatter of ≈ 0.1 dex. We fit the data to a biquadratic polynomial and obtain a result that agrees well with the observational relation(see Figure 6(c)).

The observational relations used for the subsequent bins are summarised in the table below:

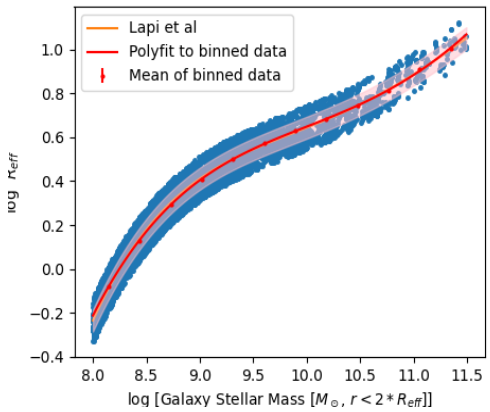
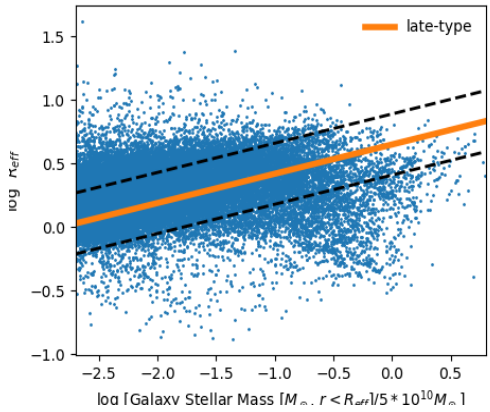


Figure 9: Mass-Size Relation bin for first z-bin

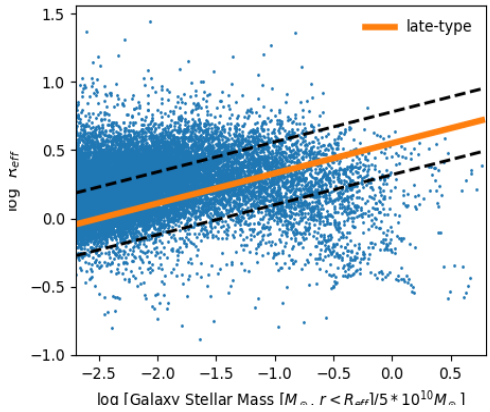
z-bin	**Reference**
0-0.05	Renzini&Peng15[13]
0.05-0.45	Speagle+14[6]
0.45-1.05	Pearson+18[21]
1.05-1.45	Kurczynski+16[22]
1.45-2.05	Kurczynski+16
2.05-2.45	Kurczynski+16
2.45-3.05	Kurczynski+16

3.1 Discrepant Mass Size Relation at high redshift

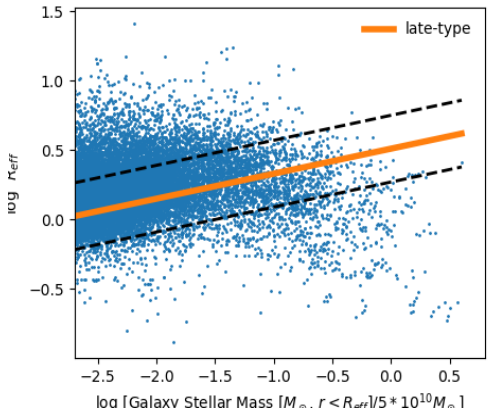
Starting from the bin $1.45 < z < 2.05$, the behaviour of simulation data at higher masses deviates from the observations of van der Wel et al and other studies of the size-mass relation. The following plots demonstrate the situation starting from this bin until $2.45 < z < 3.05$ for galaxies selected by approach 2, to convince the reader that this discrepancy is not a result of poor galaxy selection.



(a) $1.45 < z < 2.05$



(b) $2.05 < z < 2.45$



(c) $2.45 < z < 3.05$

Figure 10: Observational Mass Size Relation against TNG50 Disks

This observation has also been made in Costantin et al 2023 [23] for $z \geq 3$ but we clearly observe it for these lower redshifts as well.

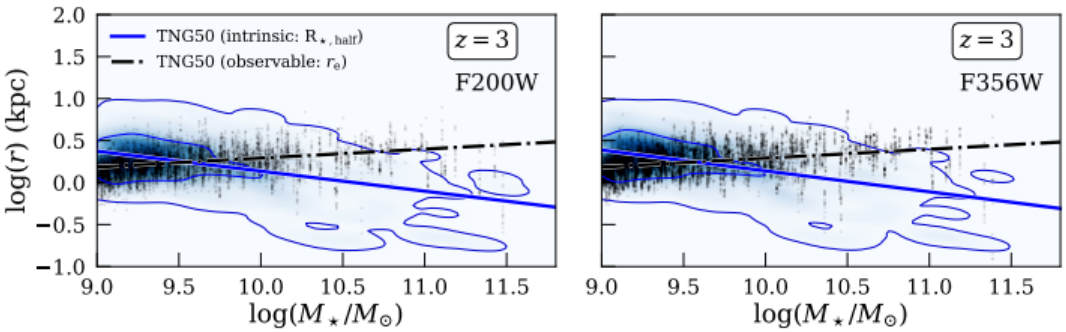


Figure 11: Mass-Size relations in TNG50 at z=3 as compared with observations from HST bands. *Source:Costantin et al 2023*

In the bin $1.45z2.05$ the slope is close to zero. For the subsequent bins it is negative. The negative slope as observed in the data at $z \geq 2$ has been predicted earlier theoretically [16]- The presence of a large number of massive and small galaxies can be attributed to extreme events of gas compaction. However this is in conflict with observations at the same redshift. Possible reasons for this disparity, as highlighted by Costantin are the following:

- High-z galaxies at the higher mass end have a more complex morphology due to which mass distribution is difficult to reproduce as a single-component model. There might be extended disky structures, biasing the light-weighted galaxy sizes to larger values than the mass-weighted ones.
- Dust attenuation of high-z galaxies may lead to obscuration of some massive and compact bulges, which underestimates the mass.
- Effective radius in observations is tracing the two-dimensional projection of the galaxy light and not its three-dimensional extension(as in TNG).

Attempts to reconcile the observations with simulations are expected to be successful after running analyses on JWST Data. The bottom line remains that the selection of galaxies via this method can give wrong results. It will exclude many more disks than include non-disks. We thus do not use the mass-size relation for these bins.

4 Results and Discussion

The resulting galaxies are stored in the form of mock catalogs for each redshift bin. These contain the following data:

Column Name	Description
z	Redshift
Snapshot	Snapshot Number
SubhaloID	
$M, M_\odot$	Total Subhalo Mass in $2R_{eff}$
$M_{gas}, M_\odot$	Subhalo Gas Mass in $2R_{eff}$
$M_{stellar}, M_\odot$	Subhalo Stellar Mass in $2R_{eff}$
$M_{DM}, M_\odot$	Subhalo Dark Matter Mass in $2R_{eff}$
$SFR(\mathrm{M}_\odot/yr)$	Star Formation Rate
V_{max}(km/s)	Maximum value of the spherically-averaged rotation curve.
R_{max}(kpc)	Distance from the galactic centre where V_{max} is attained.
σ_v(km/s)	1-D velocity dispersion of all the member particles/cells.
$R_{1/2M_*}(kpc)$	R_{eff} =Stellar Half-Mass Radius

4.1 The SFMS across redshifts

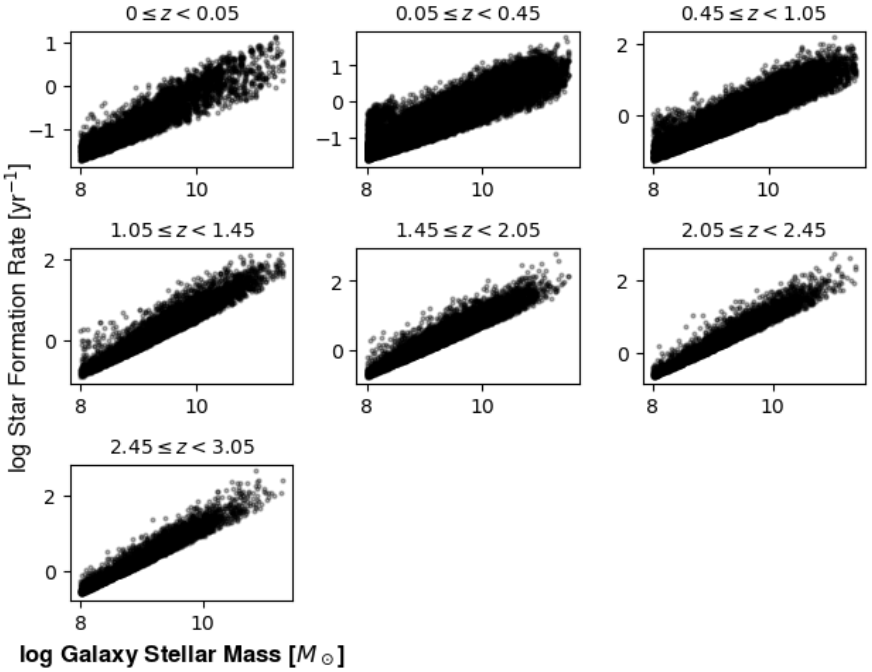


Figure 12: Star Forming Main Sequence of Selected Disks(Approach-1)

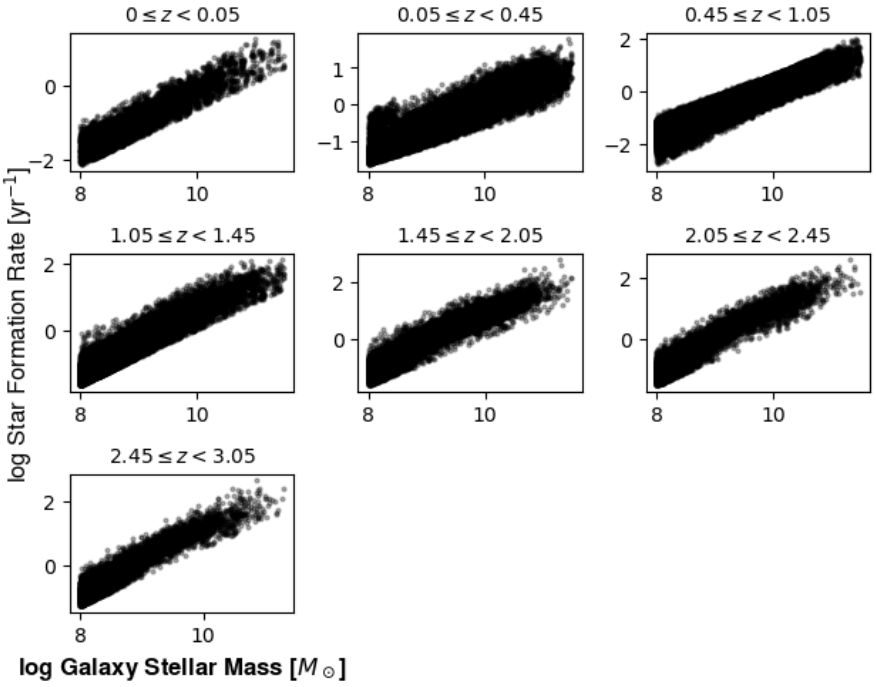


Figure 13: Star Forming Main Sequence of Selected Disks(Approach-2)

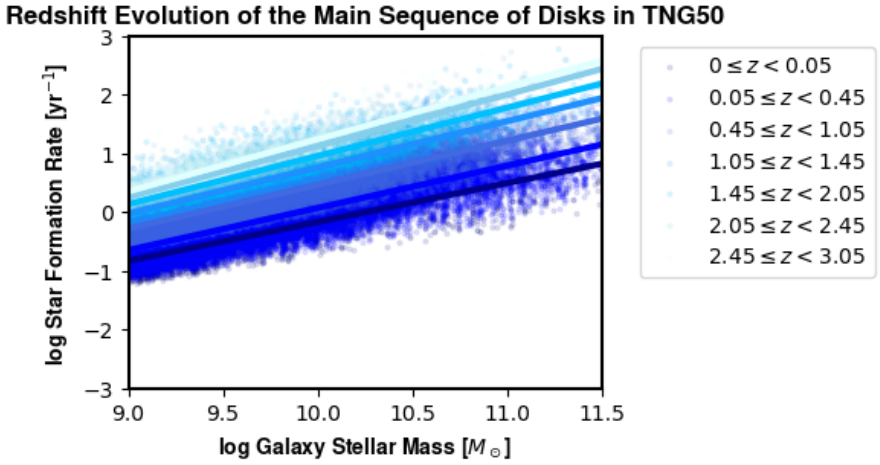


Figure 14: Redshift Evolution of MS from Disks(Approach-1)

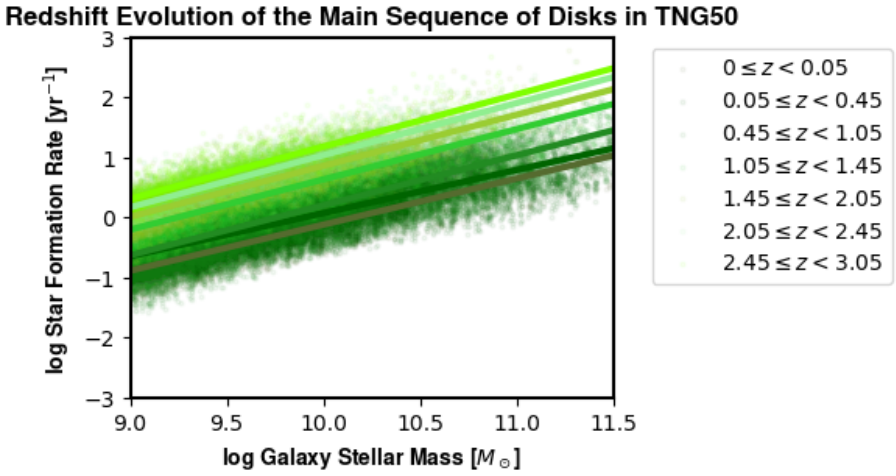


Figure 15: Redshift Evolution of MS from Disks(Approach-2)

Some elementary observations from the preceding plots and their possible explanations are as follows:

- At a given mass disks at higher redshifts had a higher SFR. This can be understood using the fact that these disks experienced rapid quenching in star formation [24], such as by running out of cold gas with time.
- There is a clear increase in scatter at the high-mass end at all redshifts. The distinction between MS, and quiescent galaxies becomes progressively less clear towards high stellar masses as they exhibit a mixture of morphologies [25]. One possible reason is that ellipticals, thought to be mergers of spirals, will tend to be more abundant at the high mass end.
- Visually there appears to be a slope evolution in the MS(increase with redshift) as reported in [6],[22] and [21]. This indicates the possibility of a mass-dependent quenching mechanism- that quenching is proportional to the stellar mass. However note that this progressive slope change is not observed in Approach-2(at the lower redshift end)
- The increase in slope with redshift also implies that galaxies become more self-similar with the increase in redshift. Massive galaxies show higher SFRs and thus a similar sSFR(specific star formation rate) in comparison to low-mass galaxies.

4.2 Size-Mass Relation across Redshifts

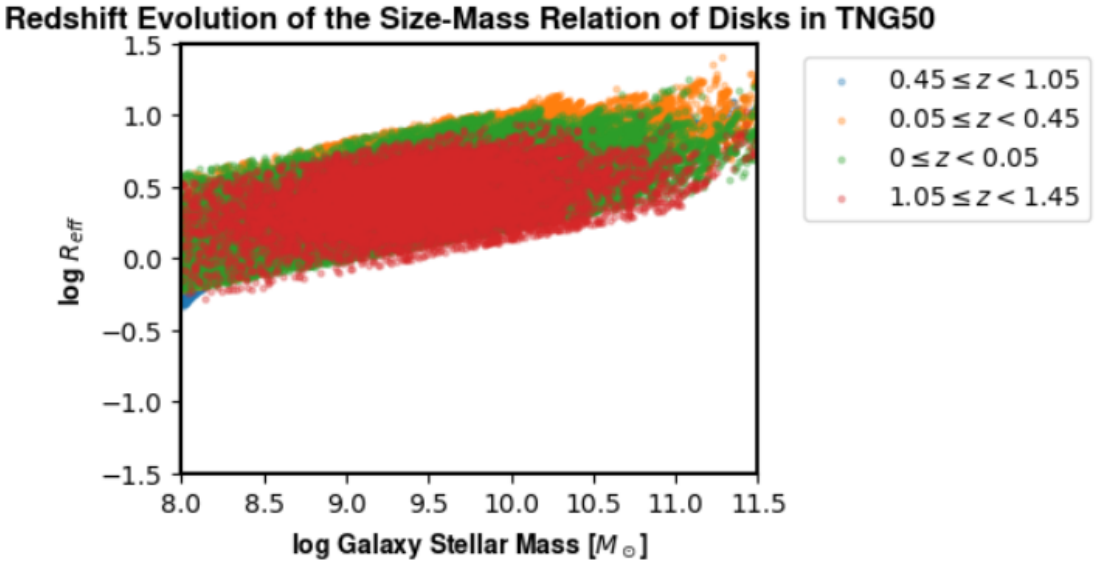


Figure 16: Evolution of the Mass-Size Relation in TNG50 across $0 \leq z \leq 3 (Approach - 1)$

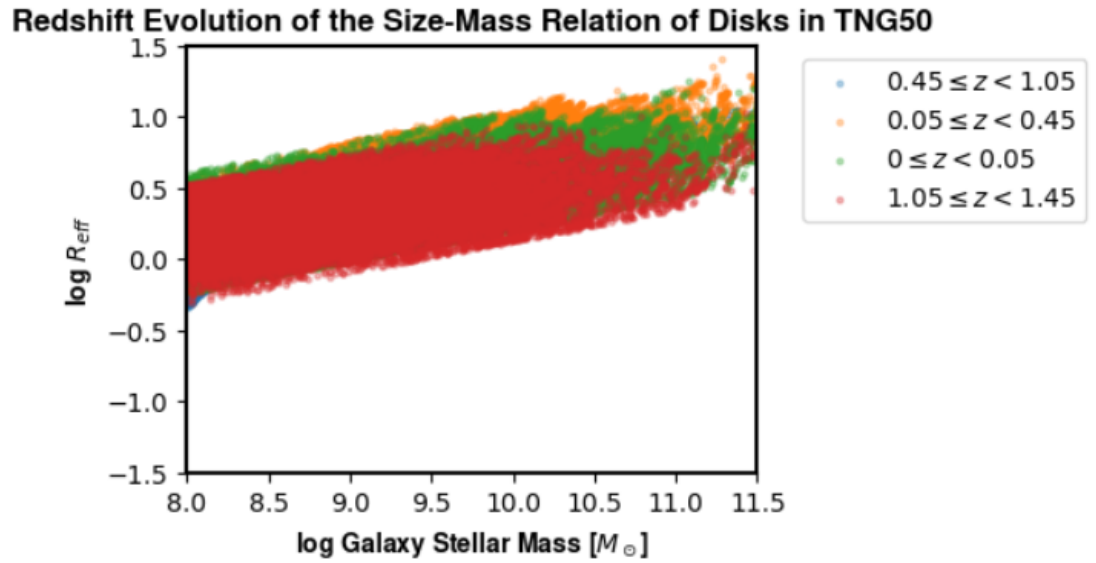


Figure 17: Evolution of the Mass-Size Relation in TNG50 across $0 \leq z \leq 3 (Approach - 2)$

The size-mass relation across redshifts as obtained from our analysis appears to have a large scatter. Nevertheless some observations can be made:

- The slope of the scaling relation remains nearly constant with redshift for masses $> 10^9 M_\odot$[18].

- The intrinsic scatter is around 0.1 dex for locals but it increases to 0.2 dex at higher redshifts. This just has to do with systematic uncertainties.

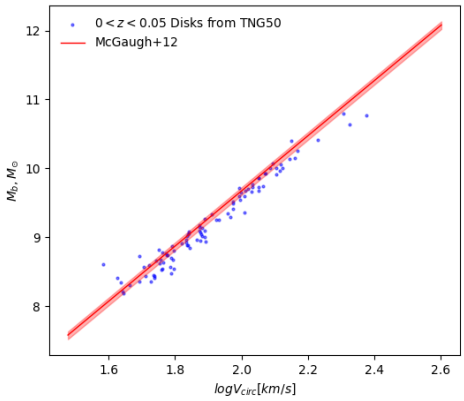
4.3 Baryonic Tully-Fisher Relation

The Tully–Fisher relation (TFR) [26] is an empirical relationship between the intrinsic luminosity(and thus M_*) of a spiral galaxy and its asymptotic rotation velocity. This is a robust correlation that has been used in cosmic distance measurements [26] and in tests of cosmology[27]. In the current context it also sets tight constraints on galaxy formation models.

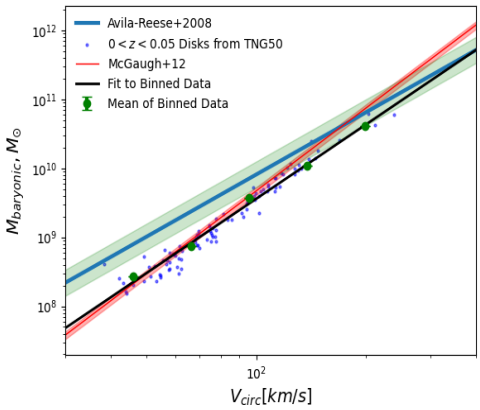
There have been different versions of this relation in literature such as by employing luminosities in various bands. One of these variants is the Baryonic Tully Fisher Relation(BTFR) when M_* is replaced by the total baryonic mass, which is the sum of stellar mass and gas mass. Doing so makes it a tighter relation. As stated by McGaugh et al 2012[27], it is

$$M_b = AV^4, A = 47 \pm 6 km^{-4} s^4.$$

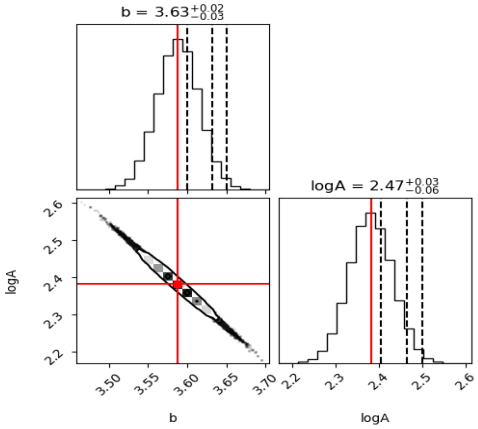
The relation is even more precise when the galaxies are gas-rich, because then the calculated baryonic mass is relatively insensitive to the Stellar Population Synthesis(SPS) Model and the wavelength-dependence of stellar mass estimates. In order to perform an independent verification of our selection of galaxies we compare this relation from our results with those of McGaugh and Avila-Reese et al 2008 [28].
To do this we draw a random sample of $\tilde{1}00$ galaxies in each catalog and plot M_b against $\log V_{circ}$ for this sample. We define $M_{baryonic} = M_* + M_{gas}$ and circular velocity $V_{circ} = \sqrt{\frac{GM_{baryonic}}{2R_{eff}}}$. Then we bin the data and fit the points to a straight line using MCMC. The figure below shows a similar set of plots for the bin $0 < z < 0.05$(all Speagle-selected).



(a) Baryonic Tully Fisher Relation for disks as compared with $0 < z < 0.05$ disks



(b) Fit to data against observations



(c) MCMC Posterior Distributions

Figure 18: A Test of selected disks using the BTFR

Although a significant fraction of galaxies lie outside the permitted scatter, the power law fit is visually not in gross disagreement with BTFR, given the mutual disagreement in observations. One thing to note is that the slope shallows with an increase in the redshift. For the highest z-bin, this discrepancy amounts to $> 4\sigma$ at the high V_{circ} end. This suggests the presence of a redshift dependent BTFR.

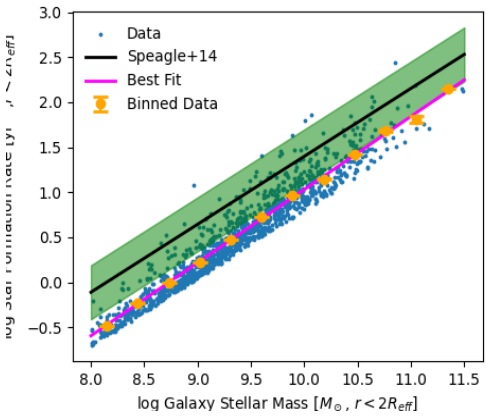
4.4 Images from the TNG Database

As another test, we generate images of these 100 $\times 7 \times 2$ samples, and manually check if they are indeed disks. The redshift bin, corresponding percentage of disks and approach are in the table.

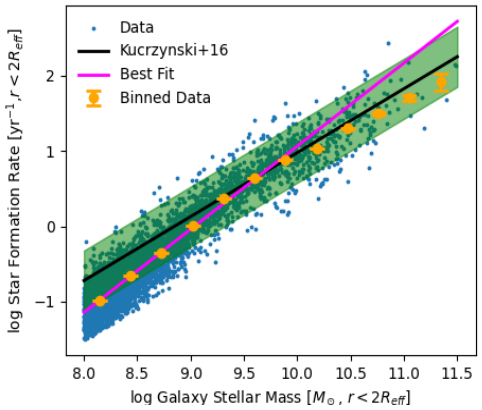
z-bin	Disk%(Approach-1)	Disk%(Approach-2)
0-0.05	95.88	97
0.05-0.45	96	98
0.45-1.05	94	98
1.05-1.45	97.97	97
1.45-2.05	88	95.96
2.05-2.45	90	89
2.45-3.05	90.91	89.90

4.5 A Potential Problem

A couple of newer papers on the Star Forming Main Sequence such as [29] state a quantitative discrepancy between results of observation and simulation. This is not surprising given the quality of fits obtained in the results,even using Approach-2(see figure below):



(a) The fit for Approach-1 (Snapshot 33) is clearly outside the permitted scatter



(b) The fit for Approach-2 in the same snapshot lies within the permitted scatter but the slope variation remains.

One way to possibly curb these effects would be to apply one single relation of the SFMS(z, $M_\star$) at median redshift z to all the redshifts within a bin. This can,in principle, be done, but would require an unreasonably high computer memory. For example, the snapshot 99 corresponding to z=0 alone bears 1013.7 GB worth data. Therefore, a more prudent way is to load the data chunk by chunk and analyze each chunk sequentially.

In this case, it might be inaccurate to entirely depend on observational results to sieve disks. A better approach would be to extract disks in TNG independently, by binning and running median estimates of SFR and then check the images of a random sample to confirm the structure. In fact, this approach was taken in one of the first TNG

papers[30]. Note that newer studies are able to quell this discrepancy[31] by employing stronger Bayesian Inference algorithms to constrain observational results. If we do take the approach of leveraging observational results, we should probably use results such as these instead of old results as in Speagle's paper. Also note that if any of this fails, we can always resort to other criteria for selecting disks such as disk-to-spheroid ratio, and the growth of stellar mass within a timescale of the order of a Hubble time. For more details see [32].

Note that it is also important to include maximum true positives than to exclude maximum false negatives- in essence, we not only need the resultant disks to be just disks, but also for them to span maximum disks in the simulation. This is why the extraction method carries a lot of importance.

5 Conclusion

The disk galaxy catalogs produced as a result of this analysis are fairly correct, as concluded from the test by BTFR and by checking random images. The BTFR in the simulated galaxies are broadly in agreement with observed results for specific ranges of circular velocity. However they differ by $> 4\sigma$ at higher V_{circ}, suggesting a z-dependent BTFR. Visualising images of a random sample of galaxies from the catalog also shows clear spiral structures in $> 85\%$ of the disks in all bins and both approaches, however on an average Approach-2 fares better according to expectations.

Our method can be used for many comparative studies with theoretical models(model validation and providing constraints for model parameters)and observations too. Some of these subjects include galaxy formation and evolution, feedback mechanisms and cosmology. However, as highlighted in the preceding section, there lies a big question as to the credibility of these results, and there is also scope for improvement in the quality of results. One can build upon our code and choose to refine the results in accordance with the demands of their research.

All the catalogs, disk IDs and scripts are available in this repository.

References

[1] P. Salucci, "The distribution of dark matter in galaxies," *The Astronomy and Astrophysics Review*, vol. 27, pp. 1–60, 2019.

[2] M. Milgrom, "On the use of galaxy rotation curves to test the modified dynamics," *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 333, Oct. 15, 1988, p. 689-693. Research supported by the Minerva Gesellschaft für die Forschung mbH.*, vol. 333, pp. 689–693, 1988.

[3] K. Masters, C. J. Lintott, and K. Schawinski, "Galaxy zoo 1: Data release of morphological classifications for nearly 900,000 galaxies," 2010.

[4] S. Wuyts, N. M. F. Schreiber, A. van der Wel, B. Magnelli, Y. Guo, R. Genzel, D. Lutz, H. Aussel, G. Barro, S. Berta, *et al.*, "Galaxy structure and mode of star formation in the sfr–mass plane from z 2.5 to z 0.1," *The Astrophysical Journal*, vol. 742, no. 2, p. 96, 2011.

[5] K. E. Whitaker, P. G. Van Dokkum, G. Brammer, and M. Franx, "The star formation mass sequence out to z= 2.5," *The Astrophysical Journal Letters*, vol. 754, no. 2, p. L29, 2012.

[6] J. S. Speagle, C. L. Steinhardt, P. L. Capak, and J. D. Silverman, "A highly consistent framework for the evolution of the star-forming "main sequence" from z 0–6," *The Astrophysical Journal Supplement Series*, vol. 214, no. 2, p. 15, 2014.

[7] K. Iyer, E. Gawiser, R. Davé, P. Davis, S. L. Finkelstein, D. Kodra, A. M. Koekemoer, P. Kurczynski, J. A. Newman, C. Pacifici, *et al.*, "The sfr–m* correlation extends to low mass at high redshift," *The Astrophysical Journal*, vol. 866, no. 2, p. 120, 2018.

[8] N. Lee, D. B. Sanders, C. M. Casey, S. Toft, N. Z. Scoville, C.-L. Hung, E. Le Floc'h, O. Ilbert, H. J. Zahid, H. Aussel, *et al.*, "A turnover in the galaxy main sequence of star formation at m* 1010 m for redshifts z¡ 1.3," *The Astrophysical Journal*, vol. 801, no. 2, p. 80, 2015.

[9] A. R. Tomczak, R. F. Quadri, K.-V. H. Tran, I. Labbé, C. M. Straatman, C. Papovich, K. Glazebrook, R. Allen, G. B. Brammer, M. Cowley, *et al.*, "The sfr–m* relation and empirical star formation histories from zfourge at 0.5¡ z¡ 4," *The Astrophysical Journal*, vol. 817, no. 2, p. 118, 2016.

[10] J. Leja, J. S. Speagle, Y.-S. Ting, B. D. Johnson, C. Conroy, K. E. Whitaker, E. J. Nelson, P. Van Dokkum, and M. Franx, "A new census of the 0.2¡ z¡ 3.0 universe. ii. the star-forming sequence," *The Astrophysical Journal*, vol. 936, no. 2, p. 165, 2022.

[11] C. Schreiber, M. Pannella, D. Elbaz, M. Béthermin, H. Inami, M. Dickinson, B. Magnelli, T. Wang, H. Aussel, E. Daddi, S. Juneau, X. Shu, M. T. Sargent, V. Buat, S. M. Faber, H. C. Ferguson, M. Giavalisco, A. M. Koekemoer, G. Magdis, G. E. Morrison, C. Papovich, P. Santini, and D. Scott, "Theherschelview of the dominant mode of galaxy growth fromz= 4 to the present day," *Astronomy amp; Astrophysics*, vol. 575, p. A74, Feb. 2015.

[12] R. Johnston, M. Vaccari, M. Jarvis, M. Smith, E. Giovannoli, B. Häußler, and M. Prescott, "The evolving relation between star formation rate and stellar mass in the video survey sincez = 3," *Monthly Notices of the Royal Astronomical Society*, vol. 453, p. 2541–2558, Aug. 2015.

[13] A. Renzini and Y.-j. Peng, "An objective definition for the main sequence of star-forming galaxies," *The Astrophysical Journal Letters*, vol. 801, no. 2, p. L29, 2015.

[14] P. Santini, A. Fontana, A. Grazian, S. Salimbeni, F. Fiore, F. Fontanot, K. Boutsia, M. Castellano, S. Cristiani, C. De Santis, *et al.*, "Star formation and mass assembly in high redshift galaxies," *Astronomy & Astrophysics*, vol. 504, no. 3, pp. 751–767, 2009.

[15] P. Santini, A. Fontana, M. Castellano, M. D. Criscienzo, E. Merlin, R. Amorin, F. Cullen, E. Daddi, M. Dickinson, J. S. Dunlop, A. Grazian, A. Lamastra, R. J. McLure, M. J. Michałowski, L. Pentericci, and X. Shu, "The star formation main sequence in the hubble space telescope frontier fields," *The Astrophysical Journal*, vol. 847, p. 76, Sept. 2017.

[16] S. Tacchella, A. Dekel, C. M. Carollo, D. Ceverino, C. DeGraf, S. Lapiner, N. Mandelker, and J. R. Primack, "Evolution of density profiles in high-zgalaxies: compaction and quenching inside-out," *Monthly Notices of the Royal Astronomical Society*, vol. 458, p. 242–263, Feb. 2016.

[17] A. Lapi, P. Salucci, and L. Danese, "Precision scaling relations for disk galaxies in the local universe," *The Astrophysical Journal*, vol. 859, no. 1, p. 2, 2018.

[18] A. van der Wel, M. Franx, P. Van Dokkum, R. Skelton, I. Momcheva, K. Whitaker, G. Brammer, E. F. Bell, H.-W. Rix, S. Wuyts, *et al.*, "3d-hst+ candels: the evolution of the galaxy size–mass distribution since z= 3," *The Astrophysical Journal*, vol. 788, no. 1, p. 28, 2014.

[19] A. Pillepich, D. Nelson, V. Springel, R. Pakmor, P. Torrey, R. Weinberger, M. Vogelsberger, F. Marinacci, S. Genel, A. van der Wel, *et al.*, "First results from the tng50 simulation: the evolution of stellar and gaseous discs across cosmic time," *Monthly Notices of the Royal Astronomical Society*, vol. 490, no. 3, pp. 3196–3233, 2019.

[20] Y.-Y. Chang, A. van der Wel, E. da Cunha, and H.-W. Rix, "Stellar masses and star formation rates for 1 m galaxies from sdss+ wise," *The Astrophysical Journal Supplement Series*, vol. 219, no. 1, p. 8, 2015.

[21] W. Pearson, L. Wang, P. Hurley, K. Małek, V. Buat, D. Burgarella, D. Farrah, S. Oliver, D. Smith, and F. van der Tak, "Main sequence of star forming galaxies beyond the herschel confusion limit," *Astronomy & Astrophysics*, vol. 615, p. A146, 2018.

[22] P. Kurczynski, E. Gawiser, V. Acquaviva, E. F. Bell, A. Dekel, D. F. De Mello, H. C. Ferguson, J. P. Gardner, N. A. Grogin, Y. Guo, *et al.*, "Evolution of intrinsic scatter in the sfr–stellar mass correlation at 0.5¡ z¡ 3," *The Astrophysical Journal Letters*, vol. 820, no. 1, p. L1, 2016.

[23] L. Costantin, P. G. Pérez-González, J. Vega-Ferrero, L. Bisigello, F. Buitrago, M. B. Bagley, N. J. Cleri, M. C. Cooper, S. L. Finkelstein, B. W. Holwerda, *et al.*, "Expectations of the size evolution of massive galaxies at 3 z 6 from the tng50 simulation: The ceers/jwst view," *The Astrophysical Journal*, vol. 946, no. 2, p. 71, 2023.

[24] O. Ilbert, H. McCracken, O. Le Fèvre, P. Capak, J. Dunlop, A. Karim, M. Renzini, K. Caputi, and S. Boissier, "Mass assembly in quiescent and star-forming galaxies since z 4 from ultravista dr1 in the cosmos field," in *SF2A-2013: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, pp. 545–548, 2013.

[25] P. Popesso, L. Morselli, A. Concas, C. Schreiber, G. Rodighiero, G. Cresci, S. Belli, O. Ilbert, G. Erfanianfar, C. Mancini, *et al.*, "The main sequence of star-forming galaxies–ii. a non-evolving slope at the high-mass end," *Monthly Notices of the Royal Astronomical Society*, vol. 490, no. 4, pp. 5285–5299, 2019.

[26] R. B. Tully and J. R. Fisher, "A new method of determining distances to galaxies," *Astronomy and Astrophysics, vol. 54, no. 3, Feb. 1977, p. 661-673.*, vol. 54, pp. 661–673, 1977.

[27] S. S. McGaugh, "The baryonic tully–fisher relation of gas-rich galaxies as a test of λcdm and mond," *The Astronomical Journal*, vol. 143, no. 2, p. 40, 2012.

[28] V. Avila-Reese, J. Zavala, C. Firmani, and H. Hernández-Toledo, "On the baryonic, stellar, and luminous scaling relations of disk galaxies," *The astronomical journal*, vol. 136, no. 3, p. 1340, 2008.

[29] P. Popesso, A. Concas, G. Cresci, S. Belli, G. Rodighiero, H. Inami, M. Dickinson, O. Ilbert, M. Pannella, and D. Elbaz, "The main sequence of star-forming galaxies across cosmic times," *Monthly Notices of the Royal Astronomical Society*, vol. 519, no. 1, pp. 1526–1544, 2023.

[30] M. Donnari, A. Pillepich, D. Nelson, M. Vogelsberger, S. Genel, R. Weinberger, F. Marinacci, V. Springel, and L. Hernquist, "The star formation activity of illustristng galaxies: main sequence, uvj diagram, quenched fractions, and systematics," *Monthly Notices of the Royal Astronomical Society*, vol. 485, p. 4817–4840, Mar. 2019.

[31] E. J. Nelson, S. Tacchella, B. Diemer, J. Leja, L. Hernquist, K. E. Whitaker, R. Weinberger, A. Pillepich, D. Nelson, B. A. Terrazas, R. Nevin, G. B. Brammer, B. Burkhart, R. K. Cochrane, P. van Dokkum, B. D. Johnson, F. Marinacci, L. Mowla, R. Pakmor, R. E. Skelton, J. Speagle, V. Springel, P. Torrey, M. Vogelsberger, and S. Wuyts, "Spatially resolved star formation and inside-out quenching in the tng50 simulation and 3d-hst observations," *Monthly Notices of the Royal Astronomical Society*, vol. 508, p. 219–235, Aug. 2021.

[32] S. Tacchella, B. Diemer, L. Hernquist, S. Genel, F. Marinacci, D. Nelson, A. Pillepich, V. Rodriguez-Gomez, L. V. Sales, V. Springel, and M. Vogelsberger, "Morphology and star formation in illustristng: the build-up of spheroids and discs," *Monthly Notices of the Royal Astronomical Society*, vol. 487, p. 5416–5440, June 2019.