Morphological classification of galaxies based on stellar kinematics in the IllustrisTNG cosmological simulation

March, 2026

Abstract

This study explores whether the stellar orbital circularity of simulated galaxies, derived from precomputed catalogs in the IllustrisTNG project, can serve as a proxy for broad morphological classification. The analysis focuses on the publicly available "Stellar Circularities, Angular Momenta, Axis Ratios" catalog, which enables a kinematic decomposition of the stellar component into disk and spheroid subsystems. Validation of this approach against the more detailed five-component kinematic decomposition in TNG50 confirms that the circularity-based disk fraction correlates most strongly with the thin disk, while the bulge fraction broadly represents the combined contribution of classical bulges and stellar halos.

The decomposition is subsequently applied to galaxies in the TNG100 simulation at redshift $z = 0$, leading to the identification of a data-motivated threshold of $F_{\rm disk} = 0.25$ to distinguish between early- and late-type galaxies (i.e., spherical/elliptical and disk-like galaxies, respectively). This threshold, being lower than the commonly adopted *ad hoc* value of 0.4 without clear physical justification, more effectively captures the diversity of disk-dominated systems and avoids the exclusion of galaxies with moderately prominent disks from the late-type morphological class. Additionally, irregular or morphologically complex systems are identified based on low total disk and spheroid mass fractions.

Utilizing this classification, a morphology-density relation is recovered that remains broadly consistent with observations: late-type galaxies dominate in the field, whereas early-type galaxies represent the most prevalent morphological type in clusters. These results demonstrate that stellar circularity alone serves as an accessible and computationally efficient morphological proxy. Finally, the potential for this classification to support machine learning efforts as a baseline or training set for future morphological studies is discussed.

1 Introduction

1.1 Historical Background

The systematic study of galaxies as autonomous stellar systems is, in cosmological terms, a relatively recent development. Throughout the 18th and 19th centuries, astronomers such as Charles Messier and William Herschel cataloged numerous diffuse, cloud-like structures known as nebulae, the fundamental nature of which remained a subject of intense debate. The central scientific question, whether these objects were local gaseous clouds within the Milky Way or distant "island universes" in their own right, was not resolved until the 1920s.

In 1920, Harlow Shapley argued that spiral nebulae were constituents of our own Galaxy, whereas Heber Curtis maintained that they were external systems of comparable scale. This controversy reached a definitive resolution when Edwin Hubble, utilizing the Hooker telescope at Mount Wilson Observatory, successfully resolved individual Cepheid variable stars within the Andromeda nebula (M31). By deriving a distance that far exceeded the established boundaries of the Milky Way, Hubble (1925, 1926) proved that the Universe is populated by a large number of galaxies. These systems are now recognized as unique, gravitationally bound assemblies of stars, interstellar gas, and dust, all situated within expansive dark matter halos (e.g., Mo et al. 2010). Characterized by significant variations in stellar mass and gas content, galaxies were initially categorized based on direct observations, with early taxonomic schemes relying primarily on prominent morphological features such as overall shape.

1.2 Morphology as a Proxy for Galaxy Evolution

The earliest classifications of galaxies emerged from direct observations, primarily grounded in morphological differences such as shape and internal structure. Although Edwin Hubble introduced the first widely adopted morphological classification scheme (Hubble 1926), shortly after it was established that galaxies are extragalactic systems distinct from the Milky Way, his system is not exhaustive. Several alternative classification frameworks have since been proposed (e.g., de Vaucouleurs 1959; Morgan & Mayall 1957; Morgan & Osterbrock 1969), yet the Hubble sequence remains the most widely used and recognized in contemporary astronomy. Despite its various subtypes, the Hubble classification broadly divides galaxies into three main morphological types: elliptical, spiral, and irregular galaxies, or, in other words, early- and late-type galaxies (where ellipticals are early-type, spirals are late-type, and irregulars are either a separate class or categorized as late-type).

The study of galaxy morphology extends far beyond the pursuit of a systematic classification scheme. Different morphological types encode crucial information about the formation history of a galaxy, and analyzing morphology can also yield insights into its future evolutionary path. Just as each individual star follows a characteristic evolutionary trajectory, so too do the structural components of galaxies (such as disks, bulges, and bars) evolve over time. These components often undergo co-evolution, interacting dynamically and chemically in ways that shape the global properties of a galaxy. Moreover, there exists a strong correlation between the morphological type of a galaxy and its physical characteristics, including color, gas content, and star formation rate, as well as the environment in which it resides (e.g., Dressler 1980a,b; Blanton & Moustakas 2009). For instance, late-type galaxies tend to be gas-rich and actively star-forming, whereas

early-type galaxies are typically red, gas-poor, and quiescent (i.e., no star formation takes place). Environmental influences, ranging from isolated field conditions to dense cluster environments, are critical in shaping these trends [1]. Consequently, studying the galaxy morphology is central to answering fundamental questions about the formation and evolution of galaxies (Conselice 2014).

1.3 Dynamical Components of Galaxies

To establish a clear conceptual framework, it is pertinent first to define the primary morphological components that are central to this analysis. The stellar component of a galaxy can be decomposed into distinct dynamical subsystems. The thin disk is a rotationally supported, flat component in which stars move on nearly circular co-planar orbits, and the thick disk is a kinematically hotter, older component with larger vertical velocity dispersions. The classical bulge forms through major mergers and resembles a small elliptical galaxy embedded in the disk center. The pseudo-bulge, despite its name, is a disk-like component that grows secularly from the disk itself through instabilities such as bars and is kinematically similar to a disk (see, e.g., Kormendy & Kennicutt 2004; Sellwood 2014). Finally, the stellar halo is a diffuse, roughly spherical envelope of old, metal-poor stars on predominantly radial orbits.

1.4 Morphological Classification in IllustrisTNG

Modern cosmological hydrodynamical simulations provide a powerful framework for testing theoretical predictions and bridging the gap between theory and observations. In the context of cosmological simulations, it is crucial to accurately reproduce observed phenomena. So far, simulations have proven to be quite useful in advancing our understanding of various topics and trends related to the evolution of galaxies and their interactions, properties, and scaling relations (e.g., Vogelsberger et al. 2020; Crain & van de Voort 2023). In this work, we will utilize the IllustrisTNG cosmological simulations[2] (Weinberger et al. 2017; Pillepich et al. 2018b; Nelson et al. 2019b). The IllustrisTNG project comprises three distinct simulation boxes, TNG50, TNG100, and TNG300, spanning a range of physical scales and mass resolutions, thereby allowing for a simultaneous study of high-resolution galaxy internal structure and large-scale cosmic environments.

When it comes to morphology, despite the richness of the data, morphological classification in simulations remains methodologically challenging. Prior efforts have primarily relied on photometric proxies through mock observations (e.g., Rodriguez-Gomez et al. 2019; Huertas-Company et al. 2019; Varma et al. 2022; Gong et al. 2025) or complex kinematic decomposition techniques, such as Gaussian mixture models applied to stellar orbits (Du et al. 2019, 2020). While these approaches offer valuable precision, they often require extensive computation or post-processing, which can limit scalability and typically represent a research study on their own. Although some of these studies have released publicly available catalogs, the datasets are often limited to one or a few specific snapshots of the cosmological simulation, focus on specific IllustrisTNG simulation boxes,

[1] The connection between the local environment and morphology of a galaxy is known as a morphology-density relation.

[2] https://www.tng-project.org/data/

or encompass a fairly conservative sample of galaxies. Hence, studies that aim to investigate specific morphological types of galaxies in more detail, or the dependence of specific global parameters (e.g., gaseous content, star formation rates, environmental proxies) on morphological types, have to either perform morphological classifications of galaxies from scratch or rely on available supplementary catalogs. The smallest IllustrisTNG simulation box, TNG50 (Nelson et al. 2019a; Pillepich et al. 2019), includes a supplementary catalog containing data on the fractional contributions of various subsystems of the stellar component of galaxies. The name of this catalog is (t) *Galaxy Morphologies (Kinematic) and Bar Properties* (Zana et al. 2022), hereafter referred to as catalog (t). The fractional contributions of components were derived using the complex kinematic decomposition tool MORDOR[3] for galaxies that have sufficient resolution (at least 1000 stellar particles) at any given snapshot of the simulation.

1.5 Kinematic Decomposition of Simulated Galaxies

In contrast to the observed galaxies, the morphology of the simulated galaxies can be determined easily via kinematic decomposition, since simulated galaxies are composed of particles, and the data include all the necessary parameters for such a procedure. The simplest kinematic decomposition is based on determining its binding energy E and the circularity of its orbit ϵ for each stellar particle in the kinematic phase space. For each stellar particle, the orbital circularity parameter ϵ is defined as the ratio $j_z/j_c(E)$, where j_z is the component of its angular momentum along the z-axis (assuming the angular momentum vector of the galaxy is aligned with the positive direction of the z-axis), and $j_c(E)$ is the maximum angular momentum a particle can have at a given energy E, assuming axial symmetry of the gravitational potential in the galactic plane[4]. The value of ϵ ranges from -1 to 1, depending on whether the rotation of the particle is opposite to (negative values) or aligned with (positive values) the overall direction of rotation in the system. The energy E represents the total energy of the stellar particle and serves as a measure of its gravitational binding to the system (assuming the total energy is negative).

The MORDOR software for the kinematic decomposition enables division of the stellar component into five subsystems based on complex criteria for ϵ and E, which was performed in the catalog (t) exclusively within the TNG50 simulation box, as mentioned previously. Although the software can be used for larger simulation boxes, this task can be resource-intensive and time-consuming. A simpler kinematic decomposition allows splitting the stellar component into only two subsystems (broadly defined and referred to as a bulge and a disk), which is even more efficient for larger samples[5]. This simpler method was applied in the supplementary catalog (c), named *Stellar Circularities, Angular Momenta, Axis Ratios* (Genel et al. 2015), hereafter referred to as catalog (c). This catalog is available for all simulation boxes and every simulation snapshot. For this reason, catalog (c) provides highly accessible information that should allow at least broad morphological classification, but we argue that it is unfortunately very underutilized.

[3] https://github.com/thanatom/mordor

[4] In other words, $j_c(E)$ corresponds to the angular momentum of a particle with the same energy E that moves on a perfectly circular orbit.

[5] This is because it does not require galaxies to have a very high particle count; decomposition can be performed even on galaxies consisting of just a few hundred particles.

1.6 Motivation and Overview

In this study, we leverage this catalog (c) to investigate whether stellar disk fractions, as defined by circularity-based decomposition, can serve as a quantitative morphological proxy for classifying galaxies into early- and late-type systems. As an exploratory extension and a certain test, we also examine whether this simple classification scheme is sufficient to recover the morphology-density relation in the simulated universe. Our goal is twofold: to assess the viability of circularity-based morphology classification as a lightweight alternative to more complex methods and to test whether IllustrisTNG reproduces a key observational trend (i.e., at least a broad agreement with the morphology-density relation) without requiring mock image synthesis, detailed orbit modeling, or any other complex method. By focusing on simplicity, we aim to lay the groundwork for more detailed follow-up studies and to make morphology analyses more accessible.

Moreover, this study constitutes a much-needed test of the precomputed data on the morphological characteristics of simulated galaxies. Despite the availability and completeness of this additional data set, the catalog (c) is, as mentioned previously, very underutilized. Most studies (e.g., Joshi et al. 2020) using the kinematic decomposition of simulated galaxies to study simply the morphology of simulated galaxies, or other relevant parameters that are sensitive to different morphological types, do so while needlessly performing the very same procedures that were used to compile the data available in the catalog (c). Such an approach is a waste of time and computational resources, as the necessary data are already publicly available in the supplementary catalog (c). Performing a similar procedure (i.e., kinematic decomposition) can only be justified if modifications or sensitivity tests are implemented, which is rarely the case. Hence, along with investigating this catalog (c) and validating it against the observational trend, the main and overarching goal of this study is not just to provide an argument for more widespread usage of this catalog, but also to recommend a practical guide on how to use it - that is, what are the physically meaningful and justified criteria for morphological classification of simulated galaxies.

This paper is organized as follows. In Section 2, we describe the simulation boxes that we use and explore the way in which we can utilize the parameters available in the catalog (c) most efficiently. In Section 3, we attempt to find a data-informed and physically meaningful criterion for the broad morphological classification of galaxies and test the classification scheme while exploring the environmental dependence of different morphological types. In Section 4, we discuss our results in the broader context of other similar and relevant studies. Finally, we give the concluding remarks in Section 5.

2 Methods

The IllustrisTNG cosmological hydrodynamical simulations of galaxy formation were carried out using the `Arepo` code (Springel 2010) and adopt the cosmological parameters from Planck Collaboration et al. (2016): a matter density of $\Omega_{\rm m} = 0.3089$, baryon density $\Omega_{\rm b} = 0.0486$, dark energy density $\Omega_\Lambda = 0.6911$, Hubble constant $H_0 = 67.74$ km s^{-1} Mpc^{-1}, power spectrum normalization $\sigma_8 = 0.8159$, and primordial spectral index $n_{\rm s} = 0.9667$. These simulations started at redshift $z = 127$, and the results are stored in 100 snapshots ranging from $z = 20$ to $z = 0$. They incorporate key astrophysical processes, including gas cooling, feedback from supernovae and active galactic nuclei (AGN), and large-scale structure formation, allowing galaxies to evolve self-consistently within large volumes and

with full cosmological context.

The simulation suite includes three flagship runs (TNG50, TNG100, and TNG300) named approximately after the side length of the simulation box (i.e., 50, 100, and 300 Mpc, respectively). In addition to differences in volume, the three boxes also differ in particle resolution, enabling studies at various scales and levels of detail. The IllustrisTNG team highlights that the TNG100 (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018a; Springel et al. 2018) offers an optimal balance between volume and resolution. The mass resolution for the TNG100 is 7.5×10^6 M$\odot$ for dark matter particles and approximately 1.4×10^6 M$\odot$ for baryonic particles (Springel & Hernquist 2003; Pillepich et al. 2018b). We will use this simulation box, since it is often viewed as the main box, and considering that it features a diversity of environments (i.e., multiple galaxy clusters and a higher number of galaxy groups) in contrast to the smaller simulation box with higher resolution (the TNG50), which contains only one Virgo-like cluster and a small number of galaxy groups. Throughout this paper, we will focus solely on the last, present-day snapshot, that is, the redshift $z = 0$.

Before we proceed with the analysis, we will first compare two supplementary catalogs, catalog (c) and catalog (t), in the TNG50 simulation box to understand and validate the physical meaning behind the parameters from catalog (c) that we will use throughout this study for morphological classification.

2.1 Validation of Mass Fractions of Different Structures from Supplementary Catalogs

Catalog (c) contains information on mass fractions of only two stellar subsystems, referred to as a bulge and a disk. A bulge is defined as the cumulative mass of all particles with $\epsilon < 0$ multiplied by two, which is noted as a common way to define a bulge. The name of this parameter in the catalog is `CircTwiceBelow0Frac`. However, as Zana et al. (2022) rightfully pointed out, this definition often does not differentiate between a bulge and a stellar halo and, without additional examination, should represent the mass fraction of both spherical components. For the disk component, the catalog contains two mass fraction parameters. In a simpler way, a disk is defined as a cumulative mass of all particles with $\epsilon > 0.7$. The second parameter that represents the mass fraction of the stellar disk is corrected by subtracting the contribution of a bulge, under the assumption that the distribution is symmetric around $\epsilon = 0$. Specifically, it is defined as the mass fraction of stellar particles with $\epsilon > 0.7$, reduced by the mass fraction of those with $\epsilon < -0.7$. The name of this second parameter is `CircAbove07MinusBelow07Frac`, and we adopt this in the present study as $F_{\rm disk}$. Based on its definition, it should represent a thin disk structure more accurately, not all disk-like components, as they can contain particles with a lower circularity. These parameters are calculated twice; one approach considers only stellar particles within 10 times the stellar half-mass radius, whereas the other takes into account all stellar particles in the subhalo. Throughout this work, we use the former. This range encompasses the majority of stellar particles, although some potential tidal features are excluded.

In contrast, catalog (t) contains information on mass fractions of as many as five different components: thin and thick disk, pseudo-bulge, classical bulge, and stellar halo. This was made possible by the elaborate analysis included in the MORDOR software and described in detail by Zana et al. (2022). The decomposition has proven remarkably successful and precise, and we will use it to compare two catalogs. We are primarily

interested in validating our initial assumptions that the bulge fractions in a catalog (c) correspond to both spherical components (i.e., bulge and stellar halo) and that the disk fraction is more in line with only a thin disk. Perhaps counter-intuitively and despite the name, pseudo-bulge is one of the disk-like components, as pseudo-bulges form from the disk itself during a secular evolution (e.g., Falcón-Barroso & Knapen 2013; Sellwood 2014) and are kinematically similar to disks (see, for example, the review by Kormendy & Kennicutt 2004).

We start by calculating Pearson correlation coefficients between the two parameters in the catalog (c) and the relevant parameters or their sum in the catalog (t). The results are shown in Fig. 1. The disk-like (t) parameter represents the sum of all disk-like structures in catalog (t): thin and thick disk, and pseudo-bulge, while the spherical (t) parameter represents the sum of the classical bulge and stellar halo.

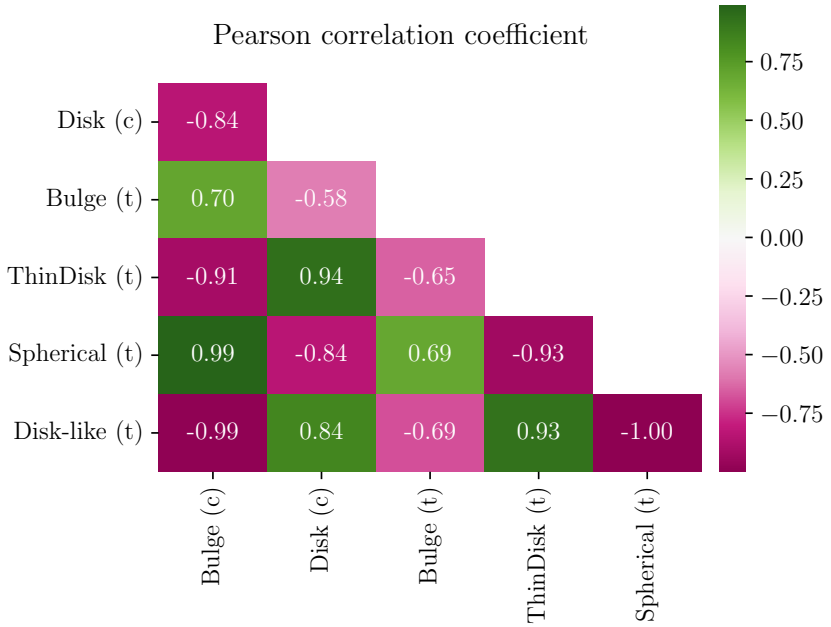

Figure 1: Correlation between fractions of different structures (or their sum) in supplementary catalogs (c) and (t), represented through the Pearson correlation coefficient. The exact values of the correlation coefficient are given in each cell.

It can be unambiguously concluded that the parameter labeled "bulge" in catalog (c) corresponds to the total mass of all spherical substructures, not just the classical bulge. This is because an analysis based solely on orbital circularity cannot distinguish between bulge and halo components, as both are characterized by nearly radial orbits, i.e., motion dominated by velocity dispersion rather than rotation. For the disk component, the distinction is less obvious. However, it can still be concluded that the parameter in catalog (c) shows a stronger correlation with the thin disk fraction than with the combined mass of all disk-like components.

Since the Pearson correlation coefficient determines the degree to which a relationship between two variables is linear, and not the actual equality between variables, we additionally examine these relationships visually. In Fig. 2 we show two-dimensional distributions of galaxies in the "Bulge (t)" – "Bulge (c)" plane (upper panel) and in the "Spherical (t)" – "Bulge (c)" plane (lower panel). The lower panel shows a clear linear relationship between the bulge component in catalog (c) and the sum of all spherical structures in

catalog (t). In contrast, the upper panel of Fig. 2 does not reveal such a linear correlation between the bulge parameters in the two catalogs, since the values of the catalog (c) are significantly higher. This discrepancy arises because the bulge in catalog (c) effectively represents the total contribution from all spherical structures, as demonstrated in the preceding analysis. The scatter in the lower panel is generally low, although at least one galaxy appears as an extreme outlier where the spherical fraction is significantly overestimated. This extreme outlier has a significant portion of its stellar mass, more than 40%, assigned to the pseudo-bulge, which is perhaps why its spherical fraction in the catalog (c) is overestimated. Other, less extreme, outliers also have somewhat prominent pseudo-bulges and typically have higher thick than thin disk fractions.

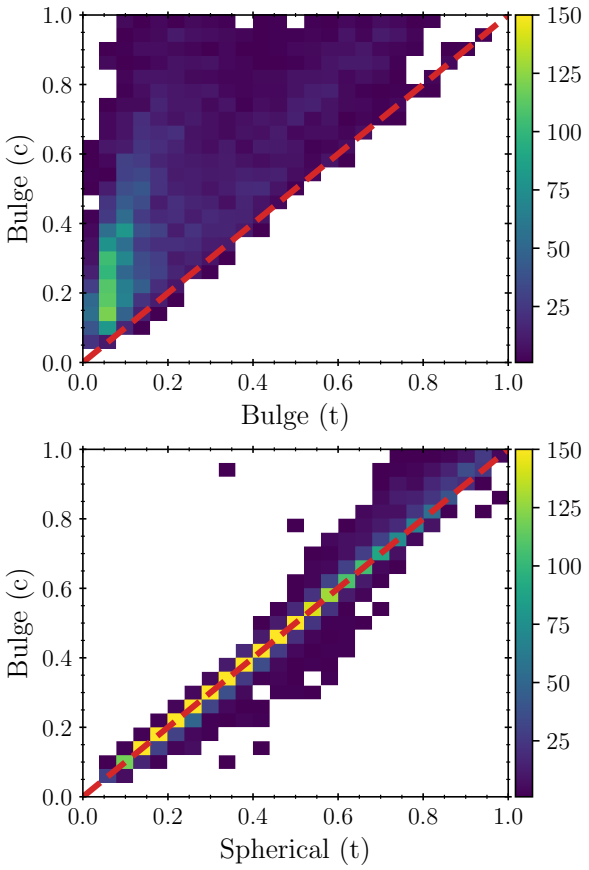

Figure 2: Correlations between bulge fraction in catalog (c) and bulge/spherical fractions in catalog (t): upper/lower panel. The red dashed line represents the $y = x$ line.

Similarly, in Fig. 3 we show two-dimensional distributions of galaxies in the "ThinDisk (t)" – "Disk (c)" plane (upper panel) and in the "Disk-like (t)" – "Disk (c)" plane (lower panel). In the upper panel, when examining the correlation between the disk component in the catalog (c) and the thin disk in the catalog (t), a clear linear trend is observed with mild scatter. In contrast, the correlation between the disk in catalog (c) and the sum of all disk-like structures from catalog (t) shows only a weak linear dependence accompanied by significant scatter. Most notably, the disk fraction in catalog (c) is almost always lower than the total fraction of disk-like structures in catalog (t). Although the disk in catalog (c) is a good proxy for the thin disk component, a scatter in the upper panel of Fig. 3 suggests that the relationship is not as tight as it is with the bulge and spherical components. However, the scatter does not appear symmetrical, and it appears that the

disk in catalog (c), when imprecise, more often underestimates the thin disk. We will consider this when we discuss the possible cut-off values of F_{disk} for the morphological classification.

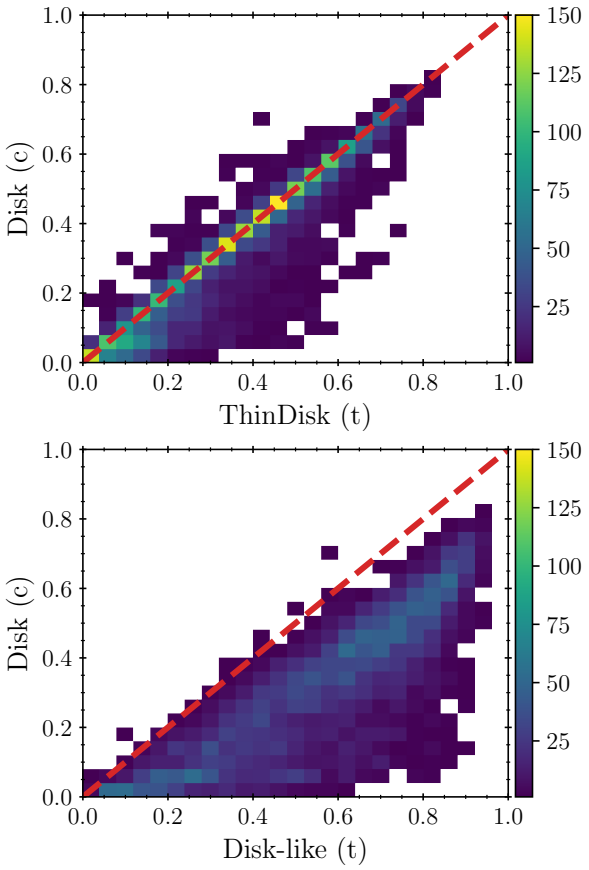

Figure 3: Similarly to Fig. 2, correlations between disk fraction in catalog (c) and thin disk/disk-like structures fractions in catalog (t): upper/lower panel. The red dashed line represents the $y = x$ line.

The main conclusion of this analysis is that it enables a clearer understanding of the parameters within the catalog (c), providing a solid foundation for the subsequent analysis to be conducted within the TNG100 simulation box. With the physical meaning of the catalog (c) parameters now established, we can approach the primary objective of this project, the morphological classification of galaxies based on the known mass fractions of different components obtained through a simple precomputed kinematic decomposition.

3 Results

Having justified the use of the catalog (c) and established the physical interpretation of the parameters `CircAbove07MinusBelowNeg07Frac` and `CircTwiceBelow0Frac`, we subsequently used these quantities as indicators of thin disk (F_{disk}) and spherical structures (F_{sph}), respectively. In principle, the combined fraction of spherical and disk-like components should not exceed unity, since by definition the fractional contribution of any component must not exceed the whole. However, there are instances where this condition is not met. In such cases, the spherical components are typically dominant and very likely

overestimated; these systems are almost certainly spheroidal or elliptical galaxies.

Conversely, there are also cases where the sum of the two components falls significantly below unity. This can be explained in two ways. First, a substantial fraction of the stellar mass may reside in substructures that are not gravitationally bound to the galaxy. Second, and perhaps a more likely explanation, a considerable number of stellar particles may be on highly eccentric orbits, which do not belong to well-defined spherical components and, due to their eccentricity, are also excluded from the disk category. This can happen, for example, if a galaxy has a significantly massive component of highly irregular shape, and perhaps a very prominent thick disk, or a strong pseudo-bulge. For this reason, it is justified to treat galaxies where the combined fraction of disk and spherical components is significantly below unity (in this study, we use a threshold of 2/3) as irregular or morphologically complex systems. When selecting this threshold, it is important to recognize that F_{disk} does not fully account for all disk-like structures and that even thin disks can often be underestimated. Therefore, adopting a slightly lower threshold is more conservative, helping to exclude galaxies with a prominent thick disk or pseudo-bulge. Additionally, decreasing this threshold would contaminate our sample of galaxies (to be classified into early- or late-type) with many genuinely irregular galaxies.

At the very beginning, we have classified a part of the sample into "irregular/complex" systems, as described above. The remaining unclassified galaxy sample may include elliptical, disk-like, and even lenticular (S0) galaxies (or early- and late-type galaxies, respectively, with S0 galaxies being an intermediate type that can be classified into both categories depending on the disk prominence). In an ideal scenario, where this sample is composed of well-separated subpopulations of elliptical and disk galaxies, the one-dimensional distribution of F_{disk} (or equivalently, F_{sph}, although this study focuses on the disk fraction) would be expected to exhibit a bimodal shape. This expectation stems from the assumption that elliptical and disk galaxies (i.e., early- and late-type galaxies) are more numerous than lenticular ones, representing a transitional morphological type. However, as shown in Fig. 4, the resulting one-dimensional distribution is not bimodal but rather monotonically decreasing.

The deviation from the ideal case is most likely due to the large number of low-stellar-mass galaxies (dwarf galaxies), which are predominantly elliptical or spheroidal. As a result, the distribution is not bimodal, making it unsuitable for determining an optimal threshold to separate early- and late-type morphological systems using this approach. This interpretation is supported by Fig. 5, which shows the two-dimensional distribution of the stellar mass versus the disk fraction, plotted as $\log(M_\star/\mathrm{M}_\odot)$ against F_{disk}. The plot reveals that galaxies with high disk fractions predominantly occupy the stellar mass range $10 \leq \log(M_\star/\mathrm{M}_\odot) \leq 11$, while a large number of early-type galaxies, with very low disk fractions, are also clearly visible, especially in the low-mass regime.

Since one approach to defining a boundary between early- and late-type galaxies involves identifying a sample whose one-dimensional F_{disk} distribution is bimodal, it is clear that the stellar mass range $10 \leq \log(M_\star/\mathrm{M}_\odot) \leq 12$ holds potential for such a distribution. Extending this interval toward lower masses would include a large number of dwarf galaxies in the sample, systems in which F_{disk} is typically low and the spheroidal component is dominant. In fact, this expectation is supported by the data shown in Fig. 6.

In Fig. 6, a clear bimodal distribution is visible, along with a local minimum located at a disk fraction of approximately 0.225. Based on this, we chose a slightly higher threshold of $F_{disk} = 0.25$, since it represents a reasonable value, at approximately a similar distance from the two peaks. Accordingly, all galaxies with a disk fraction greater than 0.25 are

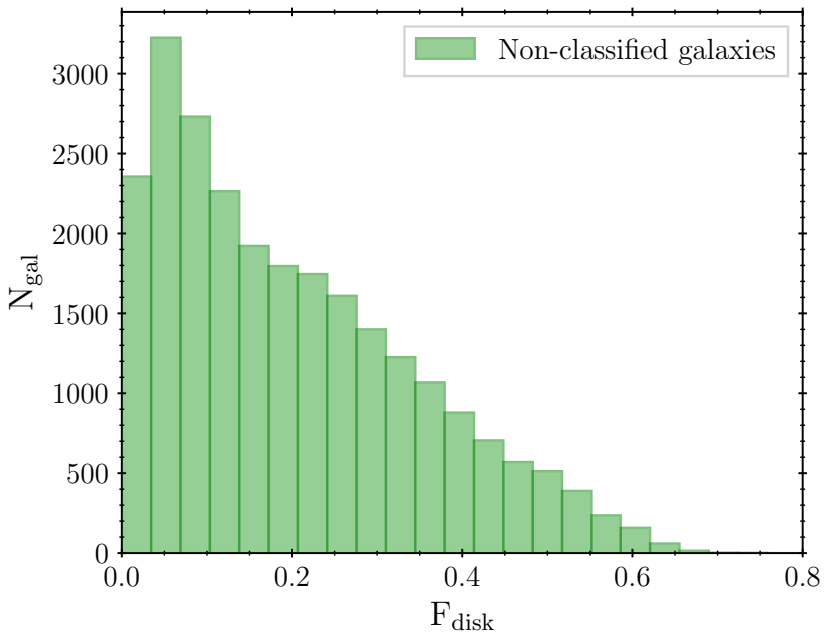

Figure 4: Distribution of disk fraction for unclassified early-type or late-type galaxies. Irregular galaxies, which were previously classified, are not included in this figure.

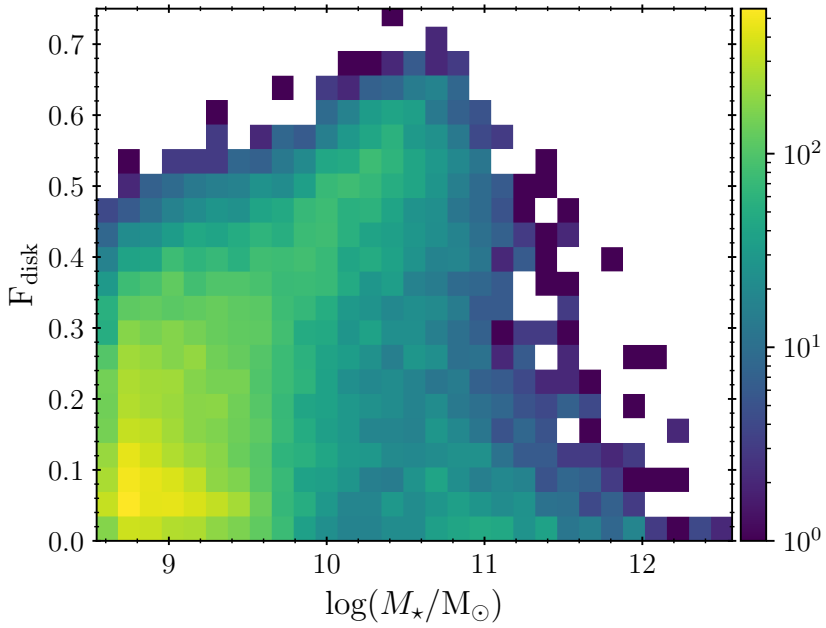

Figure 5: Diagram of disk fraction and logarithmic stellar mass, where the number of galaxies is also logarithmic (presented on the colorbar).

identified as late-type systems, while the remaining unclassified galaxies are considered to be of early morphological type. It is also worth noting that Fig. 6 allows a tentative classification of lenticular galaxies (S0). These galaxies would be characterized by F_{disk} in a certain range centered around our adopted threshold. However, such a classification was not pursued here, as this level of analysis falls outside the scope of the present study. Instead, we recognize only early- and late-type galaxies, and lenticular galaxies, given that

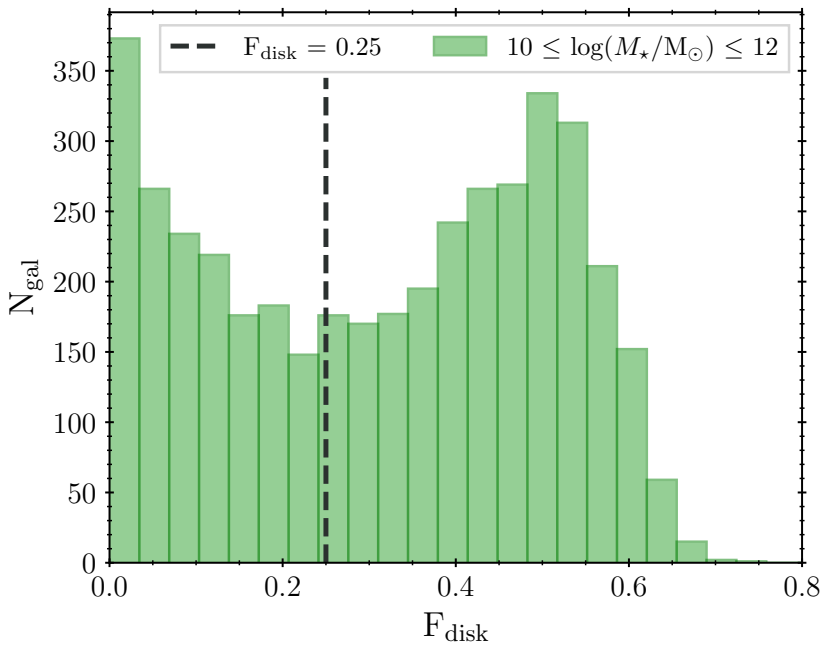

Figure 6: Disc fraction distribution for unclassified galaxies in the range $10 \leq \log(M_\star/M_\odot) \leq 12$. The black dashed line represents the cut-off value for the disk fraction, where galaxies with a disk fraction higher than the cut-off are late-type, and vice versa for early-type.

they represent a transitional morphological type, can fall into both categories, depending on how prominent their disks are.

In Fig. 7, we show the fraction of galaxies of each morphological type (irregular/complex, early-type, and late-type) as a function of stellar mass per mass bin, based on the classification scheme developed in this study. Each stellar mass bin contains galaxies normalized to unity, allowing us to compare relative abundances within bins. The plot reveals that irregular or morphologically complex galaxies (green) have a fairly symmetric distribution and occupy the intermediate mass regime, indicating a broad range of dynamical states and formation pathways. Late-type galaxies (blue) are most common in the intermediate-mass range and have a unimodal distribution that is skewed and asymmetric, featuring more low-mass galaxies than massive ones. In contrast, early-type galaxies (orange) dominate both the high-mass and low-mass ends of the distribution, having a bimodal distribution. Moreover, they are not just dominant, but perhaps the exclusive morphological type at the high-mass end. This is in agreement with observational knowledge since the most massive known galaxies are core Dominant (cD) galaxies (e.g., Morgan & Mayall 1957; Morgan & Osterbrock 1969) located at the centers of rich galaxy clusters (e.g., Matthews et al. 1964).

The trends we clearly see in Fig. 7 reflect how morphology correlates with stellar mass in the TNG100 simulation and support the rationale for selecting the intermediate-mass regime when identifying a bimodal distribution in the disk fraction.

3.1 Environmental Considerations

One way to validate the galaxy classification, and one of the goals of our study, is to examine the type of environment in which different morphological types are found. En-

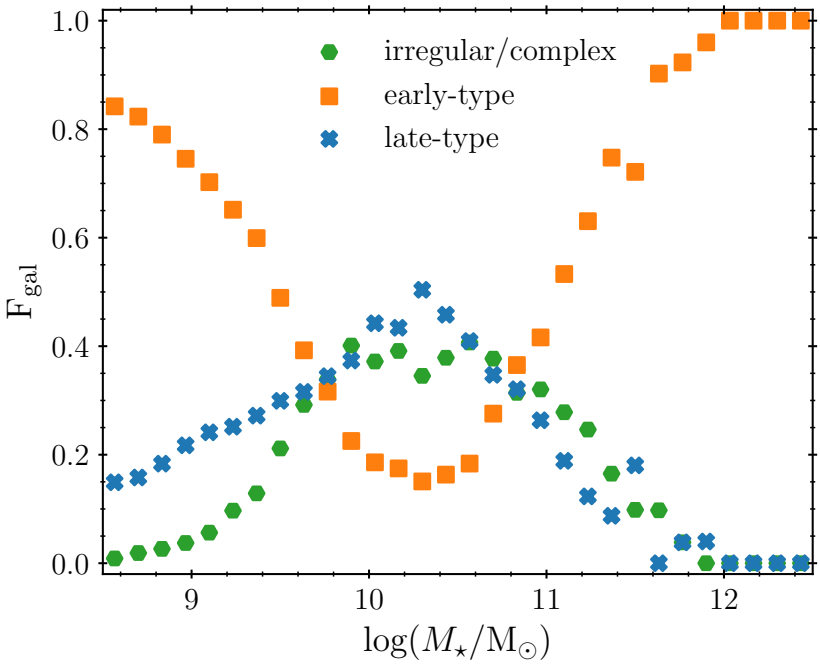

Figure 7: The fraction of galaxies per mass bin as a function of the logarithm of stellar mass. Galaxy types are color-coded: green for irregular/complex, orange for early-type, and blue for late-type galaxies (as indicated by the legend).

vironmental data are provided in the Group catalog, based on properties of identified "Friends-of-Friends" halos (that is, host halos, see e.g., More et al. 2011), and the division into clusters, groups, and field galaxies follows the criteria outlined in Paul et al. (2017). A system is classified as a cluster if the mass of its host halo exceeds 8×10^{13} $M_\odot$. Note that this applies to our study, which is focused on redshift $z = 0$, as it is not applicable at high redshifts, where proto-clusters of lower mass may exist. Field galaxies are defined as those that reside in halos with a mass below 5×10^{12} $M_\odot$ and hosting fewer than 30 subhalos, ensuring that the system is truly isolated. This additional criterion for field galaxies allows for the existence of small groups with somewhat lower halo masses but a relatively high number of subhalos, such as associations of dwarf galaxies or low-mass groups with only one fairly massive galaxy and a high number of satellites. Any galaxy not meeting the criteria for either clusters or fields is classified as residing in a group environment. This procedure constitutes a relatively coarse but practical environmental classification, providing a simple framework for investigating environmental trends. Due to contamination of the total sample with a large number of early-type galaxies at both mass ends (as we saw in Fig. 7), for environmental consideration, we focus on a subsample in the mass interval $9 \leq \log(M_\star/M_\odot) \leq 12$.

In accordance with this framework, we analyzed the disk fraction F_{disk} as a function of the type of environment, as shown in Fig. 8. It is evident that the disk fraction F_{disk} shifts toward higher values in lower-density environments, and that disk-like or late-type galaxies are practically absent in galaxy clusters. The distribution of galaxies of different morphological types across various environments was also examined. The results of this analysis are presented in Table 1.

Based on the data in Table 1, which shows the number of galaxies per morphologi-

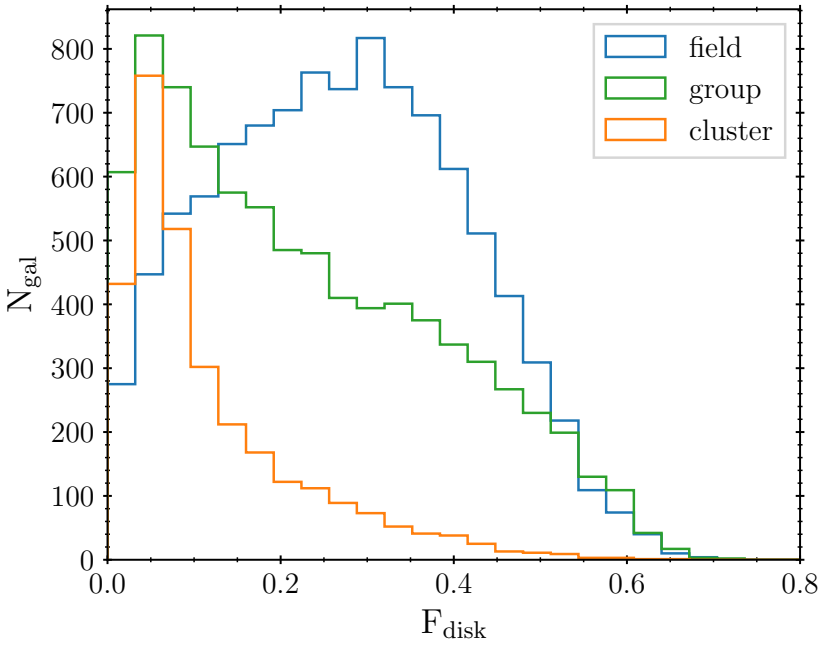

Figure 8: Distribution of disk fraction of galaxies for different environments. Due to the contamination of the total sample with a large number of early-type galaxies, a subsample in the mass interval $9 \leq \log(M_\star/\mathrm{M}_\odot) \leq 12$ was considered. Environment types are color-coded: blue for field, green for group, and orange for cluster (as indicated by the legend).

Table 1: Number of galaxies per morphological type for different types of environments. The total number of galaxies per environment type is also shown in boldface font.

	cluster	field	group
early-type	1481	4233	3553
irregular/complex	1381	1188	2440
late-type	121	4500	2140
total	**2983**	**9921**	**8133**

cal type and environment, it is fairly evident that early-type galaxies are still the most numerous. This can be explained by the fact that the redshift of the sample is $z = 0$, which corresponds to the present-day Universe. Early-type galaxies, particularly ellipticals, are expected to dominate at such low redshifts, as is reflected in the results. On the other hand, the distribution of morphological types across the three different environment categories is consistent with the previously discussed morphology–density relation. Specifically, the fraction of late-type galaxies decreases accordingly with increasing environmental density. A particularly notable trend is observed in cluster environments, where early-type and irregular/complex galaxies dominate, and late-type galaxies are nearly absent. Irregular/complex galaxies are most prevalent in groups, probably because of the more diverse and frequent interactions with other galaxies. Additionally, irregular/complex galaxies are slightly less common in the field than in groups or clusters. Even then, field irregular/complex galaxies are typically either currently interacting or have had a recent interaction with another galaxy, based on a quick, preliminary assessment of their history and nearest neighbors from subhalo and group catalogs.

Therefore, based on the combined results presented in Fig. 8 and Table 1, and the

trends that we observe with respect to different morphological types, we conclude that there is at least tentative evidence that the morphology-density relation is reproduced in IllustrisTNG.

4 Discussion

A key outcome of this study is the identification of a data-informed threshold for differentiating between early- and late-type galaxies using a simple kinematic decomposition. Using mass fractions of stellar components (estimated by using stellar circularity) from an underutilized supplementary catalog, a clear framework for morphological classification is provided that addresses existing inconsistencies in the up-to-date literature. Although prior studies have adopted similar approaches, the choice of threshold to separate disk-dominated from spheroid-dominated systems has historically lacked empirical consensus. For instance, Joshi et al. (2020) adopted a threshold of $F_{\rm disk} = 0.4$, a value that the authors acknowledged might be overly aggressive, as it potentially misclassifies galaxies with significant disk structures. Despite this, later studies frequently adopted this value, taking it at face value, without reassessing its empirical justification (e.g., Galán-de Anta et al. 2022; Łokas 2022; Fontirroig et al. 2024). Methodological divergence is further seen in the work of Park et al. (2022), who combined a lower circularity cut ($\epsilon > 0.5$) for the disk component (as opposed to the traditional value of 0.7), with a strict disk fraction threshold of 0.5.

In contrast, this analysis provides a physically motivated refinement by setting the threshold at $F_{\rm disk} = 0.25$. This value is derived directly from the bimodal distribution observed in the $10 \leq \log(M_\star/{\rm M}_\odot) \leq 12$ stellar mass range. This specific mass regime represents an evolutionary "sweet spot" where disk-dominated and spheroid-dominated galaxies coexist in appreciable numbers, allowing for a clear statistical separation. Validation against the detailed five-component decomposition in TNG50 confirms that this circularity-based disk fraction correlates most strongly with the thin disk component. A higher threshold, such as 0.4, would risk excluding transition populations, including lenticular galaxies with substantial disk components or dynamically heated spirals with pronounced thick disks, thereby underestimating the late-type fraction and biasing environmental or evolutionary interpretations.

Beyond internal consistency, tentative evidence for the reproduction of the morphology–density relation (Dressler 1980a,b; Blanton & Moustakas 2009) is found within the TNG100 simulation box. Late-type galaxies prevail in the field, whereas early-type and complex systems are more common in denser regions. This qualitative agreement with the observed morphology–density relation is encouraging, especially given the simplicity of the method. It suggests that circularity-based morphology, despite its limitations, can capture broad environmental trends and may serve as an efficient proxy for large-scale studies. Although environmental classification (field, group, cluster) remains simplified compared to continuous density measures, the emergence of these broad trends suggests that the underlying relationship between morphology and environment is robust. Importantly, while this relation was previously confirmed in EAGLE simulations (Pfeffer et al. 2023), such an examination had not been previously performed within the IllustrisTNG simulation suite. This validates the utility of catalog (c), beyond its currently limited application, for statistically robust studies without the need for computationally expensive synthetic observations.

Most importantly, this threshold serves as a versatile diagnostic tool for investigating

the co-evolution of galaxy structure and internal physics. Providing a clean separation between morphological types enables a nuanced analysis of how structure correlates with other parameters, such as the activity of the central supermassive black hole (commonly referred to as AGN). For example, establishing whether AGN feedback initiates morphological transformation or occurs as a response to structural changes requires a classification scheme sensitive to transition populations (such as lenticulars, morphologically complex galaxies, or so-called green-valley galaxies). Ultimately, the $F_{disk} = 0.25$ threshold functions as a finely tuned instrument, capturing galaxies at the precise moment of their transformation.

Furthermore, given that the catalog (c) is available across all snapshots and simulation boxes, it represents a valuable resource for future machine learning studies. The stellar circularity-based classification developed here could serve as a baseline for benchmarking or even as a training set for supervised learning models aimed at predicting galaxy morphology from other observable or derived properties. As upcoming observational campaigns produce an overwhelming influx of data, the development of reliable automated classification pipelines will become strictly mandatory. The physically motivated threshold of $F_{disk} = 0.25$ provides a clean, pre-validated baseline that is crucial for training these future artificial intelligence models, ensuring that they are grounded in dynamical reality rather than arbitrary cutoffs. Such applications could help bridge the gap between simulation-based classifications and those used in observational surveys, facilitating domain adaptation in a cosmological context.

5 Conclusion

Understanding the morphological diversity of galaxies remains one of the most significant challenges in modern extragalactic astronomy. In observational studies, classification is frequently constrained by projection effects, limited resolution, and the high computational cost of complex image processing. This study addresses these challenges by leveraging cosmological magneto-hydrodynamical simulations from the IllustrisTNG project. Unlike isolated simulations, which model galaxies in an idealized vacuum, the cosmological framework allows galaxies to evolve self-consistently within large volumes, shaped by their environment, mergers, and internal feedback processes.

This investigation established that stellar orbital circularities from the underutilized catalog (c) provide a robust basis for broad morphological classification. Validation against high-resolution kinematic decomposition in TNG50 confirmed that these circularity-based fractions accurately represent physically motivated structural components, such as the thin disk and combined spherical subsystems. A primary contribution of this work is the identification and justification of a data-motivated threshold of $F_{\rm disk} = 0.25$ to differentiate between early- and late-type galaxies (i.e., spherical/elliptical and disk-like galaxies, respectively). Derived directly from the bimodal distribution observed in the intermediate stellar mass range, this threshold captures the structural diversity of galaxies more effectively than traditional, more aggressive thresholds and mitigates systematic biases in identifying late-type systems.

The successful recovery of the morphology–density relation within IllustrisTNG, which remains broadly consistent with observational trends, further validates the reliability and efficiency of this classification scheme. Beyond simple categorization, the $F_{\rm disk} = 0.25$ threshold functions as a finely tuned instrument for identifying galaxies during critical

stages of transformation, providing a versatile diagnostic tool for investigating the co-evolution of galaxy structure and internal physics. This precision is essential for future studies aiming to isolate, for example, the role of AGN feedback in initiating morphological change versus responding to it. Given its availability across all snapshots and volumes, this classification represents a valuable resource for machine learning applications, providing a scalable baseline to bridge the gap between theoretical models and observational surveys. Ultimately, by replacing computationally prohibitive classification methods with an accessible and efficient proxy, this approach eliminates a major methodological bottleneck. This framework allows the astronomical community to shift its focus and computational power away from the labor of classifying galaxies and toward understanding the fundamental physical drivers behind their cosmic evolution.

References

Blanton, M. R. & Moustakas, J. 2009, ARA&A, 47, 159
Conselice, C. J. 2014, ARA&A, 52, 291
Crain, R. A. & van de Voort, F. 2023, ARA&A, 61, 473
de Vaucouleurs, G. 1959, Handbuch der Physik, 53, 275
Dressler, A. 1980a, ApJS, 42, 565
Dressler, A. 1980b, ApJ, 236, 351
Du, M., Ho, L. C., Debattista, V. P., et al. 2020, ApJ, 895, 139
Du, M., Ho, L. C., Zhao, D., et al. 2019, ApJ, 884, 129
Falcón-Barroso, J. & Knapen, J. H. 2013, Secular Evolution of Galaxies (Cambridge University Press)
Fontirroig, V., Gomez, F. A., Jaque Arancibia, M., Dolfi, A., & Monsalves, N. 2024, arXiv e-prints, arXiv:2411.19723
Galán-de Anta, P. M., Sarzi, M., Pillepich, A., et al. 2022, MNRAS, 517, 5992
Genel, S., Fall, S. M., Hernquist, L., et al. 2015, ApJ, 804, L40
Gong, J.-Y., Lin, W., Tang, L., & Lan, Y. 2025, arXiv e-prints, arXiv:2504.18042
Hubble, E. P. 1925, The Observatory, 48, 139
Hubble, E. P. 1926, ApJ, 64, 321
Huertas-Company, M., Rodriguez-Gomez, V., Nelson, D., et al. 2019, MNRAS, 489, 1859
Joshi, G. D., Pillepich, A., Nelson, D., et al. 2020, MNRAS, 496, 2673
Kormendy, J. & Kennicutt, Robert C., J. 2004, ARA&A, 42, 603
Łokas, E. L. 2022, A&A, 662, A53
Marinacci, F., Vogelsberger, M., Pakmor, R., et al. 2018, MNRAS, 480, 5113
Matthews, T. A., Morgan, W. W., & Schmidt, M. 1964, ApJ, 140, 35
Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution (Cambridge University Press)
More, S., Kravtsov, A. V., Dalal, N., & Gottlöber, S. 2011, ApJS, 195, 4
Morgan, W. W. & Mayall, N. U. 1957, PASP, 69, 291
Morgan, W. W. & Osterbrock, D. E. 1969, AJ, 74, 515
Naiman, J. P., Pillepich, A., Springel, V., et al. 2018, MNRAS, 477, 1206
Nelson, D., Pillepich, A., Springel, V., et al. 2019a, MNRAS, 490, 3234
Nelson, D., Pillepich, A., Springel, V., et al. 2018, MNRAS, 475, 624
Nelson, D., Springel, V., Pillepich, A., et al. 2019b, Computational Astrophysics and Cosmology, 6, 2
Park, M., Tacchella, S., Nelson, E. J., et al. 2022, MNRAS, 515, 213

Paul, S., John, R. S., Gupta, P., & Kumar, H. 2017, MNRAS, 471, 2
Pfeffer, J., Cavanagh, M. K., Bekki, K., et al. 2023, MNRAS, 518, 5260
Pillepich, A., Nelson, D., Hernquist, L., et al. 2018a, MNRAS, 475, 648
Pillepich, A., Nelson, D., Springel, V., et al. 2019, MNRAS, 490, 3196
Pillepich, A., Springel, V., Nelson, D., et al. 2018b, MNRAS, 473, 4077
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13
Rodriguez-Gomez, V., Snyder, G. F., Lotz, J. M., et al. 2019, MNRAS, 483, 4140
Sellwood, J. A. 2014, Reviews of Modern Physics, 86, 1
Springel, V. 2010, MNRAS, 401, 791
Springel, V. & Hernquist, L. 2003, MNRAS, 339, 289
Springel, V., Pakmor, R., Pillepich, A., et al. 2018, MNRAS, 475, 676
Varma, S., Huertas-Company, M., Pillepich, A., et al. 2022, MNRAS, 509, 2654
Vogelsberger, M., Marinacci, F., Torrey, P., & Puchwein, E. 2020, Nature Reviews Physics, 2, 42
Weinberger, R., Springel, V., Hernquist, L., et al. 2017, MNRAS, 465, 3291
Zana, T., Lupi, A., Bonetti, M., et al. 2022, MNRAS, 515, 1524