Laidlaw Research Project

The Cosmic Waltz: Galaxies, their Circumgalactic Medium Halos, and the Cosmic Web

Supervised by:

Dr Rita Tojeiro

School of Physics and Astronomy, University of St Andrews

By Var Konovalova

University of St Andrews

Essay outline

1. Abstract
2. Introduction
   - A. Cosmic web
   - B. CGM
3. Research aim
4. Hypothesis
5. Method
   - A. Source of data
   - B. Profiles
     - a. Density
     - b. Volume-weighted thermal pressure
     - c. Mass-weighted temperature
     - d. Mass-weighted metallicity
   - C. CGM boundary
   - D. Properties
   - E. Histogram split
     - a. 3D plot
   - F. Changing of perspectives
   - G. DBScan split
   - H. Variation of initial conditions
     - a. Omega_m
     - b. Sigma 8
6. Results
   - A. Comparison of median-split and density-split profiles
   - B. Mass-weighted temperature
   - C. Mass-weighted metallicity
   - D. Interpreting the property set
   - E. Overview of the 4 profiles
   - F. Variation of cosmological parameters
7. Discussion
8. Research challenges
9. Impact and next steps
10. Conclusion
11. Acknowledgements
12. Bibliography

1. Abstract

The circumgalactic medium (CGM) is a diffuse halo of gas surrounding galaxies that acts as both a reservoir of star-forming fuel and a channel for feedback-driven outflows. Despite its importance, the role of the CGM in linking galaxies to their large-scale environment, the cosmic web, remains poorly understood. This project investigates whether the properties of the CGM are measurably influenced by cosmic environment using data from the CAMELS IllustrisTNG simulations.

A sample of approximately 450 galaxies with available radial profiles of density, thermal pressure, temperature, and metallicity was analysed. Galaxies were classified using two complementary approaches: (i) median-split populations based on intrinsic properties, including stellar mass, gas metallicity, specific star formation rate (sSFR), and velocity; and (ii) density-based spatial clustering (DBSCAN) to approximate cosmic web structures. For each classification, average CGM profiles were constructed and compared to assess correlations between galactic properties, environmental context, and CGM behaviour.

The analysis revealed limited direct correlations between environment and CGM properties. While spatial clustering did not produce strong distinctions in CGM profiles, property-based classifications indicated notable trends. Galaxies with higher stellar mass and metallicity displayed systematically hotter CGM temperatures, consistent with enhanced radiative feedback and enriched outflows. By contrast, systems with higher sSFR showed cooler CGM profiles, reflecting the requirement of cold gas for efficient star formation. Velocity trends were less robust, suggesting that peculiar motions may not strongly regulate CGM thermodynamics.

These findings underscore the CGM as an active agent in galaxy evolution, regulating inflows and outflows, even if large-scale environment leaves only subtle signatures. Future work will extend the dataset to include the full simulated galaxy population and compare these results to upcoming observational data from facilities such as the Vera C. Rubin Observatory. This approach will refine our understanding of how galaxies and the cosmic web co-evolve.

2. Introduction:

A. Cosmic web

The Cosmic web was discovered in 2008 by the University of Colorado at Boulder [1]. In 1929, Edwin Hubble et al. published a revolutionary paper following the observation that galaxies, visible from the Milky Way, were mostly moving away from us [2]. Another peculiarity was that the further one looked, the faster those galaxies were receding [2]. In 2011, two independent groups were awarded the Nobel prize in physics for ‘the discovery of the accelerating expansion of the universe’, back in 1998 [3]. This discovery provided evidence for vacuum energy, which works in opposition to gravity. It “pushes” things apart. Since then, the composition of the universe has been refined and quantified to be about 30% matter and about 70% vacuum energy today [4].

Until the early years of the 1920s, our galaxy (the Milky Way) was thought to be the largest structure in the universe and thought to contain all the stars in our night sky [5]. In 1923, Edwin Hubble definitively proved the existence of other galaxies, outside of the Milky Way [6]. Later in 1982, we were able to zoom out and discover larger structures, bigger than just a galaxy alone [7]. The universe is made up of a lot of gas, stars, dust, and galaxies. All these together, we had assumed should be rather homogeneously distributed, but as more and more galaxies were catalogued and plotted, a peculiar pattern emerged. It looked like an even larger structure existed: the cosmic web. In the cosmic web, high density regions are called nodes, slightly less density regions connecting nodes are called filaments, and low-density regions are called voids [8]. These 3 types of regions make up the cosmic web, 'the large-scale backbone of the universe' [9].

This large structure, the environment that galaxies exist in, is relatively not well understood but assumed to influence galaxy properties. The degree to which the cosmic web influences galaxies is a hot area of study.

B. CGM

The definition of a galaxy is also difficult to nail down and write, such that every expert would agree. The difficulty is in the fact that there is no distinct border, so it is difficult to define where a galaxy ends and begins. Galaxies tend to have a decreasing density profile as a function of radius from the centre. The centre of a galaxy is a high-density region, usually with a super massive black hole [10], lot of stars, and lots of hot gas. The further you go from the centre, the less stars and gas you will find. This distribution is typically found to decrease about exponentially. An issue that was encountered in the 1970s was that the rotational speeds of objects in the outskirts of mapped galaxies were too fast [16]. The galaxies should have flown apart, but as they do not, this indicates that there must be some mass accelerating this beyond what we are able to see. This unseen mass is called dark matter and adds a further level of difficulty in defining the borders of a galaxy.

A transitional area between the inter-stellar medium (ISM) and inter-galactic medium (IGM) is called the circumgalactic medium (CGM). The ISM is what is between stars and the IGM is the matter between galaxies, so the rough definition of the CGM used for this research is that it is the matter around galaxies but not between galaxies.

It appears that the cosmic web has a lot less of an effect on the ISM than the IGM. Studying the CGM allows us to connect the large scale of the cosmic web with the smaller scale of the galaxy and the objects (stars, dust, and gas) within them.

This research will classify the sampled galaxies into two or three populations, then plot various profiles of these populations to quantify any differences or similarities. The reason for the classification is to attempt to uncover any underlying distinctions between these galaxies in the cosmic web that would help inform on the cosmic web itself.

3. Research aim:

To study and quantify the influence of the cosmic web on the CGM of galaxies.

4. Hypothesis:

The environment that a galaxy exists in is expected to influence its chemical composition, as well as its mass and type. It is likewise expected that a galaxy's properties and ISM within it, can influence its environment and surrounding IGM.

The hypothesis is that different correlations will be found for the different populations, and that different galaxy populations will favour different regions of density in the cosmic web.

5. Method:

A. Source of data

The relationship between galaxy properties, CGM profiles, and the galaxy environment in our universe is investigated using CAMELS IllustisTNG simulation.

CAMELS stands for Cosmology and Astrophysics with MachinE Learning Simulations. It is a project that employs various simulation types and methods to simulate our universe since the Big Bang with snapshots taken at even intervals, to allow various type of studies and investigations to uncover the secrets of our universe. CAMELS is designed to bridge cosmology and astrophysics in a large dataset to make use of machine learning techniques [14]. These 2 branches of astronomy are tightly interlaced; however, the theoretical models for each branch are only accurate at vastly different scales, complicating the study of the universe.

The IllustrisTNG project specifically aims to better understand the physical processes driving galaxy formation and evolution. Each CAMELS IllustrisTNG 1P (one parameter) simulation used in this study has slightly different initial conditions, only varying one parameter per simulation; they have been allowed to run until the present day or that simulations equivalent. As all data is simulated, it allows for intrinsic properties of galaxies to be accurately measured, which is not easily done from observational data.

B. Profiles

A CAMELS IllustrisTNG 1P simulation, at time set to the present day, has about 20,000 galaxies in the last snapshot of the simulation, of which about 450 galaxies have had

4 different profiles plotted as a function of radius from the greatest density region, the galaxy's centre. The profiles are recorded at 25 radial values, which are on a logarithmic scale from about 10^{-1} kpc to about 10^{5} kpc. The profiles plotted are:

a. Density

The density profile represents the radial distribution of gas mass per unit volume in the CGM. It traces the overall structure and concentration of material surrounding the galaxy.

b. Volume-weighted thermal pressure

The volume-weighted thermal pressure profile describes the pressure exerted by gas particles, averaged over volume. It traces the dynamical support of the CGM against gravitational collapse.

c. Mass-weighted temperature

The mass-weighted temperature profile measures the thermal state of the gas, averaged with respect to mass. It traces the energy content of the CGM, and highlights regions influenced by heating and cooling processes.

d. Mass-weighted metallicity

The mass-weighted metallicity profile represents the abundance of heavy elements relative to hydrogen and helium, averaged with respect to gas mass. It traces the chemical enrichment of the CGM through stellar evolution and feedback.

C. CGM boundary

The range of the CGM is defined by the virial radius: the radius within which the scatter of the measured velocities of the particles are at a maximum [11]. It can also be defined as being the radius from the centre of an over density, within which the density is 200 times that of the critical density (a fundamental, calculated value). Within this radius, the system is in virial equilibrium: the sum of the time-averaged potential energy and two times the time-averaged kinetic energy equals zero [12]. As the equilibrium is difficult to measure the over density definition is the one used in this study, which is why the plots show virial radius as $R_{200}c$. The inner limit on the CGM is taken to be 10% [13] of $R_{200}c$ and the outer limit as 100%.

D. Properties

Each galaxy in the dataset is associated with roughly thirty measurable properties, each represented by a single value per galaxy. From these, four key properties were selected for this study. Stellar mass quantifies the total mass contained in stars and provides a measure of the galaxy's growth and evolutionary state. Gas metallicity describes the abundance of elements heavier than helium in the interstellar and circumgalactic gas, tracing chemical enrichment from stellar processes. Specific star formation rate (sSFR) measures the rate of ongoing star formation relative to stellar mass, characterising the galaxy's current level of star-forming activity. Galaxy centre-of-mass velocity represents the bulk motion of the system, capturing its kinematic state within the larger cosmic environment.

The selection of 4 properties from around 30, was done via visual inspection for the most different looking graphs.

E. Histogram split

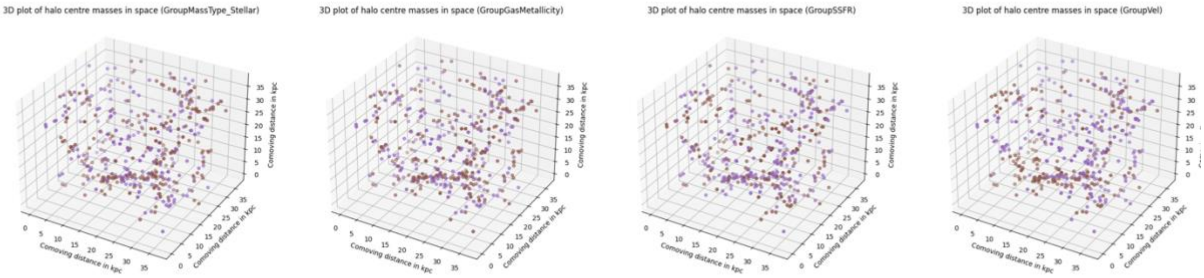
To assess how variations in each property influenced the CGM profiles, a histogram of the ~450 galaxy values was first constructed. The median of each property distribution was then identified, and the sample was split into two subsets: galaxies with values above the median and those below. This approach ensured two balanced populations of approximately 225 galaxies each, enabling a consistent comparison. Alternative measures of central tendency, such as the mode or median-based subdivisions, were tested but produced discontinuities in the resulting profiles; for this reason, the mean was adopted to generate smoother composite profiles.

For each subset, the four CGM profiles were averaged and plotted together to compare the two galaxy populations defined by the property in question. Error bars were added of the standard deviation calculated for every radial point, from the halos being averaged. The expectation was that one mean profile would reveal a positive correlation while the other would show either a weaker or negative trend. On linear axes, these correlations were not evident, and when the axes were transformed to log–log scale, the interpretation remained unclear. To address this, the profiles were instead plotted with a logarithmic radial axis, which provided a clearer representation of structural differences across scales.

The results indicated that the density and thermal pressure profiles showed little variation between the two populations, whereas the temperature and metallicity profiles exhibited more noticeable differences, though their overall correlations remained similar.

a. 3D plots

To complement the radial profile analysis, the galaxy sample was also examined in three-dimensional space. Using the same histogram-based median split, the centres of mass of all halos were plotted in comoving coordinates and colour-coded according to their property classification (above or below the median). This approach provided a spatial view of the distribution of halos for each property, enabling a visual search for large-scale patterns or clustering associated with stellar mass, gas metallicity, specific star formation rate, and velocity. No strong spatial trends or systematic differences between the two populations were evident in these plots.



[Figure 6]. Three-dimensional spatial distribution of halo centres coloured by property classification above and below the median. Each panel shows galaxies split by stellar mass, gas metallicity, specific star formation rate, and peculiar velocity, respectively. The plots provide a visual test for large-scale clustering differences between populations; no clear spatial segregation or preference for particular environments is apparent.

F. Changing of perspective

The decision was made to introduce a density-based classification, as the averaged profiles had shown only limited differences and the three-dimensional spatial plots revealed no clear preference for one galaxy subset to cluster or occupy a distinct space. While this outcome was unexpected, it may be attributed to the relatively small sample size, and further considerations are addressed in the Discussion section. Overall, the distributions appeared approximately homogeneous, despite the expectation that distinct clustering features of the cosmic web might emerge.

The motivation for reclassifying the galaxies in this way was to approach the research question from the opposite perspective. In the initial part of the analysis, galaxy properties served as the independent variable, with the CGM structure treated as the dependent variable. This allowed for the investigation of the CGM primarily from the ISM and effects from inside the galaxy, a bottom-up approach. In the following section, the density environment is instead taken as the independent variable, with the CGM profile trends considered as the dependent variable. This allowed for an investigation of the CGM primarily for the IGM and other effects from outside the galaxy, a top-down approach.

G. DBScan split

A density-based clustering approach was also applied using the DBSCAN algorithm on the three-dimensional positions of ~450 galaxy halos. The algorithm was implemented with parameters ε = 4 Mpc, defining the maximum distance between two points for them to be considered neighbours, and min_samples = 5, defining the minimum number of points required to form a cluster. This procedure identified several clusters, with the two largest and most populated assigned the status of “nodes,” following their approximate analogy to dense regions of the cosmic web. The remaining clusters were grouped into a second category labelled “medium density,” while galaxies that did not meet the clustering criteria were treated as a third category representing the least dense environments, “lowest density”.

This classification therefore produced three subsets: nodes, medium-density clusters, and lowest-density (noise) galaxies. Only the first category draws a direct analogy to a cosmic web structure, as the other two do not strictly correspond to the established definitions of filaments and voids, thus only one carries an analogous name. For each subset, the four CGM profiles were averaged to compare environmental trends. Unlike the earlier property-based splits, this approach produced three curves on the profile plots rather than two. The resulting profiles, however, did not exhibit the degree of variation that might have been expected; possible explanations for this outcome are explored in the Discussion section.

H. Variation of initial conditions

In addition to property-based and environment-based classifications, variations in the initial conditions of the simulations were examined. Several cosmological parameters can be adjusted to alter the starting conditions of a simulated universe; here, the focus is placed on two fundamental parameters, Ωm and σ_8. Each parameter sets the conditions under which the simulation evolves, and although their values remain fixed throughout a given run, they determine how the system develops over time. To investigate their influence, a series of simulations were analysed in which only one parameter was varied while all others were kept at their fiducial values. The same histogram-based property splits and density-based splits used earlier were repeated for these simulations to enable consistent comparison of CGM profiles.

a. Omega_m

Ωm is the matter density parameter, representing the normalised contribution of matter to the total energy density of the universe. Varying Ωm effectively changes the matter content at the beginning of the simulation and thus modifies the initial conditions from which structure forms. In this study, Ωm was varied from 0.1 to 0.5 in steps of 0.1, spanning a range around the fiducial cosmological model. The currently accepted model of a flat universe requires that the total sum of matter, vacuum energy, and radiation energy densities equals one; adjusting Ωm while holding the others fixed isolates the role of matter density in shaping cosmic structure.

The same histogram split and density split analyses were performed on five differently simulated universes with varying Ωm. Although the expectation was that differences in matter density would imprint noticeable changes on the CGM profiles, the results revealed relatively little variation between the simulations. This outcome was somewhat unexpected and is considered further in the Discussion section.

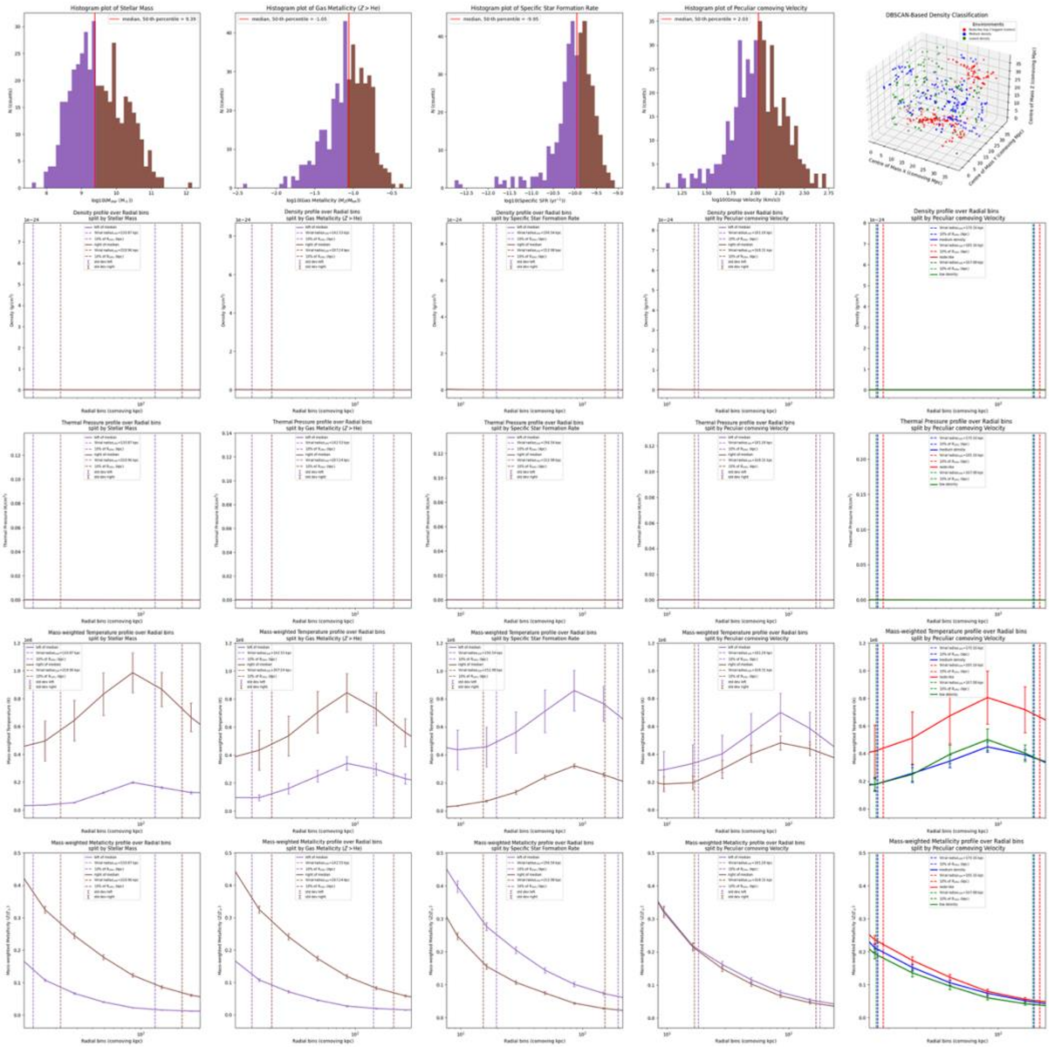
b. Sigma 8

σ_8 is a parameter that quantifies the amplitude of matter density fluctuations on scales of 8 Mpc/h, expressed as the root-mean-square (RMS) of the smoothed density field [14]. It serves as a measure of clustering strength, sometimes described qualitatively as the “clumpiness” of the universe. Unlike Ωm, σ_8 is not tied to a single constituent of the universe but instead encapsulates the statistical variation of the matter density field. In these simulations, σ_8 was varied from 0.6 to 1.0 in intervals of 0.1, with all other cosmological parameters fixed to their fiducial values.

Repeating the same analyses as for Ωm, the results showed little distinction between the different universes simulated with varying σ_8. This aligns with the outcome of the Ωm tests, as the strength of the matter density fluctuations encoded by σ_8 is intrinsically linked to the matter density itself. The limited variation observed is therefore consistent with expectations and is discussed in more detail in the Discussion section.

6. Results

Properties such as the star formation rate (SFR) remain central to galaxy classification and interpretation. In the simulations used here, SFR values are explicitly tracked to account for stellar evolution and gas recycling, ensuring that galaxies can form new stars as older populations die. Although in principle SFR might seem straightforward to measure observationally, in practice it is complicated by the long lifetimes of stars and the extended timescales for gas collapse. Because direct comparisons of stellar light across millions or billions of years are impossible, indirect methods are typically required, each with its own uncertainties. For this reason, simulations provide a crucial complement to observations, offering controlled environments for studying galaxy properties across scales ranging from the large-scale cosmic web down to atomic processes.

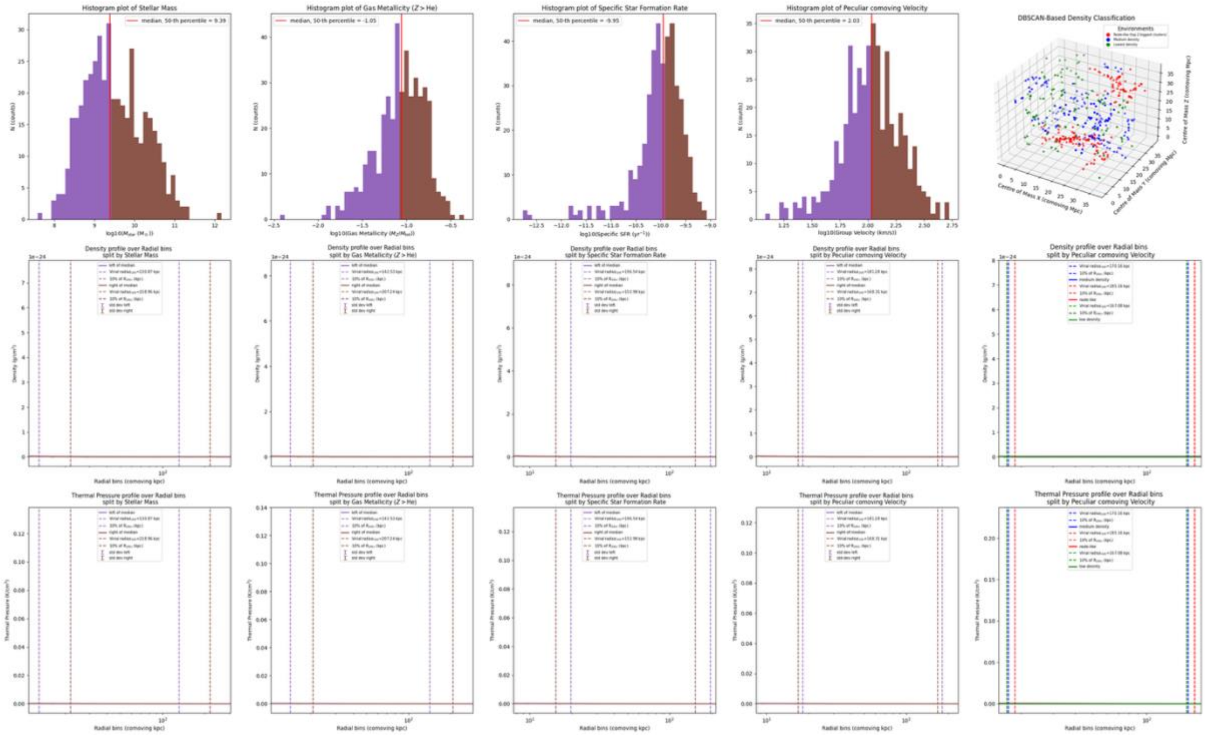


[Figure 1]. Histograms of galaxy properties (stellar mass, gas metallicity, specific star formation rate, and peculiar velocity) with median splits, together with the DBSCAN-based density classification. The rows below show the corresponding averaged CGM profiles: density (second row), thermal pressure (third row), mass-weighted temperature (fourth row), and mass-weighted metallicity (fifth row). Each profile is plotted as a function of radius, with galaxies separated by property value or environmental density. This figure provides an overview of the four diagnostics used throughout the analysis and illustrates the relative sensitivity of each to both property-based and density-based classifications.

A. Comparison of median-split and density-split profiles

When comparing the four median-split averaged profiles to those classified by environmental density, larger differences appear in the property-based splits than in the density-based classification. This outcome is somewhat unexpected, since regions of the cosmic web are thought to exhibit distinct populations shaped by their chemical enrichment and dynamical histories. One possible explanation is that cosmic web environments cannot be uniquely defined by density alone, and thus the adopted classification may not accurately capture the underlying physical structures.

The density profile declines with radius, consistent with theoretical expectations. Within the virial radial range defined by the 10% and 100% boundary lines, the average density decreases systematically outward, reflecting the dilution of baryons in the CGM. The thermal pressure profile shows a similar decline with radius, peaking in the inner bins before falling off at larger distances. Both density and pressure therefore provide consistent evidence that the CGM becomes more diffuse and less dynamically supported with increasing distance from the galaxy centre.



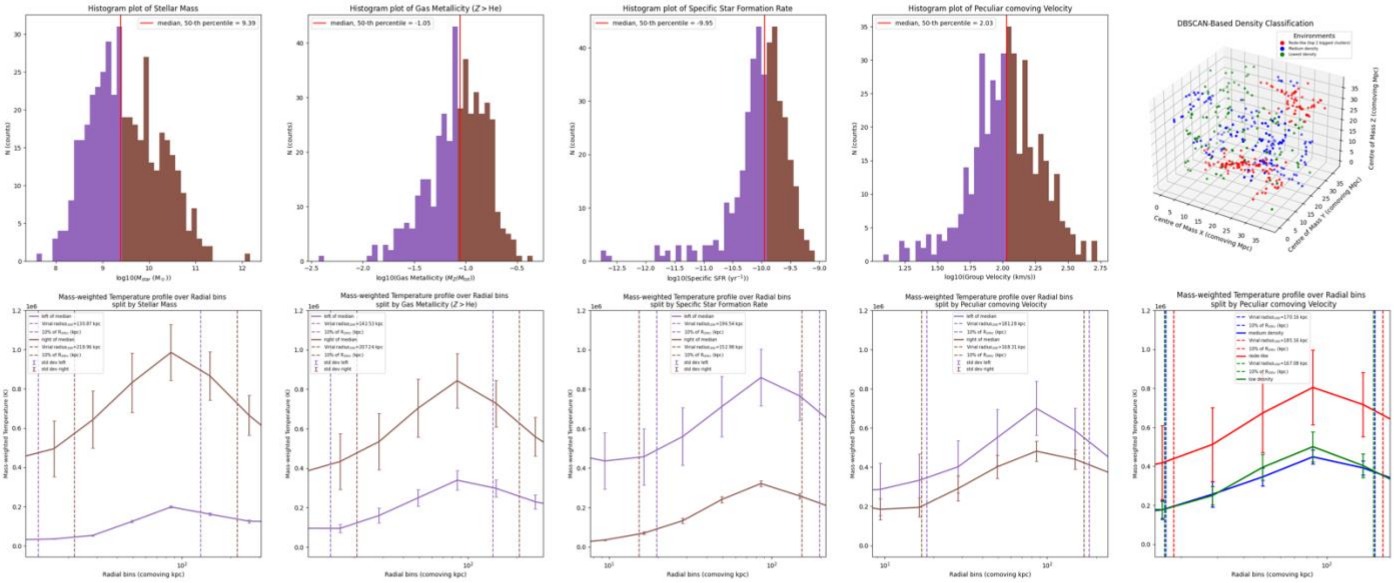
[Figure 2]. Histograms of galaxy properties (stellar mass, gas metallicity, specific star formation rate, and peculiar velocity) used for the median-split classification, alongside the DBSCAN-based density classification in three-dimensional space. Below, the corresponding averaged profiles of gas density (middle row) and thermal pressure (bottom row) are shown for each split. The density and pressure profiles exhibit very little variation between the classified populations, with values remaining nearly flat across the radial range, suggesting that these diagnostics are less sensitive to differences in either galaxy properties or environmental density.

B. Mass-weighted temperature

The mass-weighted temperature profile reveals clearer correlations with galaxy properties. Halos with higher stellar mass exhibit hotter CGM temperatures, likely reflecting the increased energy input from more numerous stars and associated feedback processes. A similar trend is seen for halos with higher gas metallicity, consistent with the role of supernovae and stellar winds in both enriching and heating surrounding gas.

By contrast, halos with higher specific star formation rates (sSFR) show systematically cooler CGM temperatures across all radial bins compared to those with lower sSFR. This anti-correlation can be explained by the need for cold gas in star formation: when abundant cold gas is available, stars can form efficiently, whereas hotter gas cannot cool and collapse as easily. Peculiar velocity shows a similar anti-correlation, with faster-moving halos exhibiting cooler CGM temperatures on average.

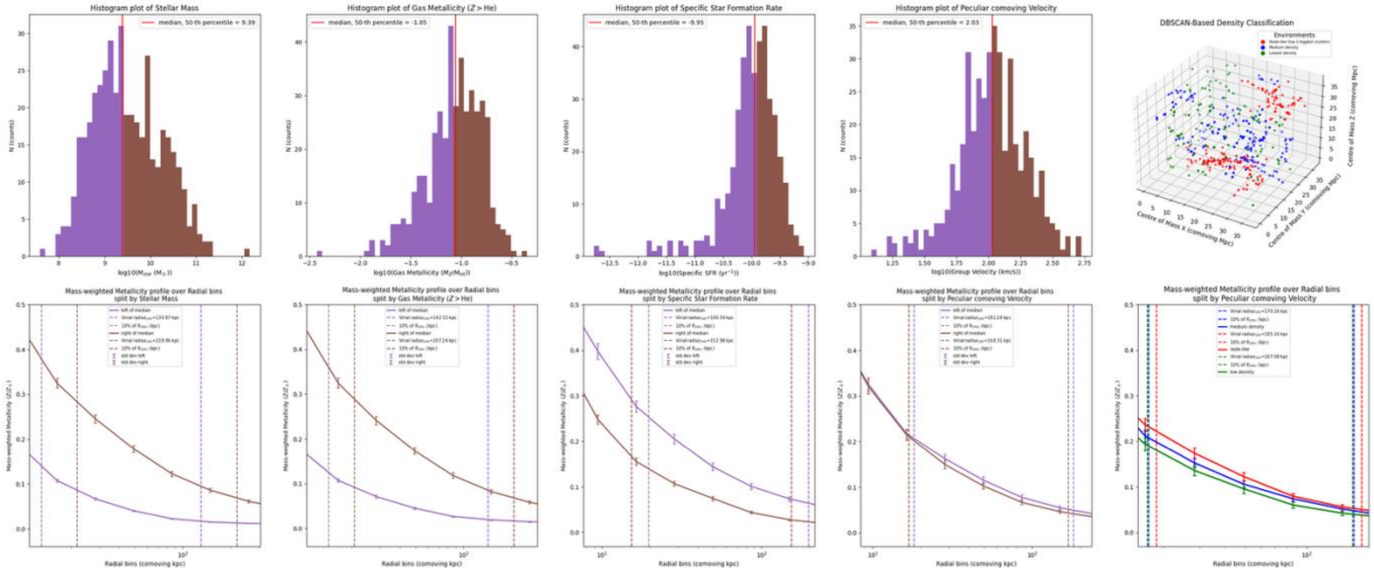
An additional feature appears in the velocity-temperature relation: halos with the largest peculiar velocities show elevated temperatures at larger radii. This behaviour is difficult to interpret physically and may reflect a statistical correlation rather than direct causation. A similar caution applies to the metallicity-velocity relation, where outer radial bins again show unexpectedly high temperatures.



[Figure 3]. Histograms of galaxy properties (stellar mass, gas metallicity, specific star formation rate, and peculiar velocity) with median splits, together with the DBSCAN-based density classification. The lower panels show mass-weighted temperature profiles as a function of radius for each classification scheme. Clear correlations are observed for stellar mass and metallicity, where higher values correspond to hotter CGM temperatures, while inverse correlations are found for sSFR and peculiar velocity. Error bars indicate the spread within each radial bin.

C. Mass-weighted metallicity

The metallicity profiles follow a declining trend with radius, consistent across all classification schemes. Galaxies with higher stellar mass and metallicity show higher CGM enrichment in the inner regions, reflecting the cumulative output of star formation and feedback. Conversely, galaxies with higher sSFR or peculiar velocities exhibit somewhat lower metallicity values, suggesting that ongoing accretion of pristine gas or dynamical disturbances dilute chemical abundances in the CGM.



[Figure 4]. Histograms of galaxy properties (stellar mass, gas metallicity, specific star formation rate, and peculiar velocity) with median splits, together with the DBSCAN-based density classification. The lower panels show mass-weighted metallicity profiles as a function of radius for each classification scheme. In all cases, metallicity decreases with radius, consistent with expectations of enriched inner regions and more pristine outer gas. Higher stellar mass and metallicity galaxies show enhanced CGM enrichment, while galaxies with higher sSFR or peculiar velocities display relatively lower metallicities, particularly at small radii.

D. Interpreting the property set

The investigated properties-stellar mass, gas metallicity, sSFR, and peculiar velocity-capture complementary aspects of galaxy and CGM evolution. Stellar mass and sSFR primarily reflect intrinsic galaxy processes, although both can be influenced by environmental conditions. Gas metallicity straddles internal and external influences: it depends on enrichment mechanisms within galaxies while also reflecting inflows and outflows that link to environment. Peculiar velocity is more strongly tied to external conditions, as large-scale gravitational interactions shape the overall motion of galaxies. This mix of internally and externally influenced parameters allows a broader assessment of the processes governing CGM structure.

E. Overview of the four profiles

The four profiles studied: density, thermal pressure, mass-weighted temperature, and mass-weighted metallicity, each provide distinct insights. The density profile declines with radius, as expected from baryonic distribution in halos. The thermal pressure profile mirrors this decline, tracing the dynamical support of gas. The mass-weighted temperature profile integrates density and pressure information, highlighting the role of feedback and cooling, while the metallicity profile quantifies chemical enrichment and its radial distribution. Taken together, these diagnostics provide a multifaceted view of the CGM.

F. Variation of cosmological parameters

Finally, simulations with varied cosmological parameters Ωm and σ_8 were analysed. Neither set of variations produced significant differences in the inter-profile relations when compared to the fiducial simulation, which best represents current cosmological constraints. This null result suggests that, within the tested ranges, CGM profiles are relatively insensitive to moderate changes in these parameters.

An animation of the variation of cosmological parameters reflected in the summary of the above explored graphs can be found here:
https://tinyurl.com/TNG_Omegam_&_sigma8

7. Discussion

Working with simulated data requires caution when translating findings to real observations, yet simulations also provide invaluable opportunities for analysis. Many physical quantities are extremely difficult or impossible to measure directly in observations, but simulations allow such quantities to be extracted, and their influence studied across a wide range of scales, from the formation of stars and dust to the cosmic web itself. At the same time, simulations carry intrinsic limitations, particularly in terms of resolution and computational demands. It is not feasible to simulate every particle in the universe across its 14-billion-year history; instead, stars, black holes, and gas are approximated as particles whose gravitational interactions reproduce large-scale structure.

Splitting galaxies into two or three arrays simplifies the methodology, but averaging across ~225 galaxies without outlier rejection means that peculiar cases may disproportionately affect the results. Furthermore, the dataset analysed here includes only ~450 galaxies out of roughly 200,000 simulated objects, since only these had available profiles. A more complete dataset covering all galaxies and halos would allow more robust classification into cosmic web structures and reduce sensitivity to sampling effects.

Resolution and sample size also shape the reliability of conclusions. Higher resolution simulations, or simulations run at multiple cosmic times, could provide more realistic representations of galaxy and CGM evolution. While large-scale structures can be approximated with coarse “particles,” the underlying physics is governed by processes at much smaller, subatomic scales that cannot be captured in full detail.

The analysis itself was limited to four CGM profiles: density, thermal pressure, temperature, and metallicity. While these profiles offer important insights, incorporating a broader range of diagnostics could uncover additional relationships not visible here. An alternative classification approach was also tested, splitting galaxies below the 10th percentile and above the 90th percentile. However, this produced results similar to the median split, and subtracting the two sets of profiles introduced discontinuities, especially at very small and very large radii, with little effect in the main CGM region (10–100% of the virial radius). This region, which forms the focus of this study, is widely but not universally accepted as the boundary between galaxies and the IGM

Methodological decisions were made with the aim of aligning with established practice. For instance, the virial radius was used to define CGM boundaries, since it is readily available in the simulations. An alternative boundary, the half-light radius (the radius enclosing half of a galaxy’s luminosity), is often cited as a measure of galaxy extent, but could not be used here as it is not provided in the data. The absence of such alternative measures highlights the constraints imposed by simulation outputs.

8. Research challenges

This project faced several challenges that shaped both the methodology and interpretation of results. A key limitation was the reliance on simulated rather than observational data. While simulations allow for precise measurements of quantities that are otherwise difficult to obtain, they also rely on assumptions about physics and cosmology, and are limited by computational resolution. Another challenge was the relatively small subset of galaxies with available CGM profiles (about 450 out of 20,000 in the simulation), which constrained the statistical power of the analysis. Methods such as median and percentile splits, as well as clustering algorithms, sometimes produced null or inconclusive results. These challenges highlight the complexity of studying structures at the intersection between galaxies and the cosmic web.

9. Impact and next steps

The findings of this research emphasise the importance of the circumgalactic medium as an active component of galaxy evolution, even when clear environmental correlations are difficult to detect. Future work will involve expanding the dataset to include more galaxies, refining classification methods to better capture cosmic web structures, and comparing simulation results to upcoming observational data, such as that expected from the Vera C. Rubin Observatory. This research has already provided opportunities for dissemination through the Laidlaw Scholars Conference and may form the basis for further presentations or publication.

10. Conclusion:

This meta-analysis of the CAMELS Illustris-TNG 1P dataset did not uncover strong causal links between galaxy properties and CGM profiles, nor did it confirm the expected correlations. Nonetheless, the investigation highlights areas where further work could be highly valuable. Despite the predominance of null results across both linear and logarithmic analyses, the methodology developed, and the code implemented provide a framework for future studies.

Specifically, the relationships between stellar mass, gas metallicity, specific star formation rate, and peculiar velocity with respect to the four CGM diagnostics density, thermal pressure, temperature, and metallicity did not reveal consistent or causal trends. However, the approaches tested here demonstrate how simulations can be used to structure such analyses and point to areas where larger datasets, higher resolution, and alternative boundary definitions may yield new insights into the interplay between galaxies, their CGM, and the cosmic web.

11. Acknowledgements

This project would not have been possible without the guidance of my supervisor, Dr Rita Tojeiro, whose advice and steady encouragement shaped both the direction and the confidence behind this work. I am also deeply thankful to Finlay and Celina for their constant support, and to my dear friend Len Prabhakar, who helped me make sense of it all when the data felt overwhelming. I gratefully acknowledge Lord Laidlaw for making this opportunity possible through the Laidlaw Scholars Programme, which has allowed me to pursue research while developing as a leader. Finally, I hope this essay speaks not only to astronomy but to the wider community, as a reminder that science thrives when curiosity and connection cross disciplines.

12. Bibliography

1. University of Colorado at Boulder. Cosmic web discovered. ScienceDaily [Internet]. 2008 May 20 [cited 2025 Aug 27]. Available from: https://www.sciencedaily.com/releases/2008/05/080520152013.htm
2. Choi C. Edwin Hubble's expanding universe. Live Science [Internet]. 2018 [cited 2025 Aug 27]. Available from: https://www.livescience.com/64527-edwin-hubble-universe-expanding.html
3. Nobel Prize Outreach. The Nobel Prize in Physics 2011 [Internet]. 2011 [cited 2025 Aug 27]. Available from: https://www.nobelprize.org/prizes/physics/2011/summary/
4. ESA/Hubble. Composition of the universe. ESA/Hubble [Internet]. [cited 2025 Aug 27]. Available from: https://esahubble.org/science/composition_of_universe/
5. Space.com. Milky Way Galaxy: Facts About Our Galactic Home. Space.com [Internet]. 2017 Mar 21 [cited 2025 Aug 27]. Available from: http://www.space.com/19915-milky-way-galaxy.html
6. Evans B. The Great Debate – 100 years later. Astronomy.com [Internet]. 2020 Apr 25 [cited 2025 Aug 27]. Available from: https://astronomy.com/news/2020/04/the-great-debate-of-shapley-and-curtis–100-years-later
7. Webster A. The clustering of quasars from an objective-prism survey. Mon Not R Astron Soc [Internet]. 1982;199(3):683–705. doi:10.1093/mnras/199.3.683. Available from: https://doi.org/10.1093/mnras/199.3.683
8. Massara E, et al. The clustering of galaxies in the cosmic web. Mon Not R Astron Soc [Internet]. 2024;528(2):3186–3204. Available from: https://academic.oup.com/mnras/article/528/2/3186/7596451
9. NASA. Mapping the Cosmic Web. NASA Science [Internet]. [cited 2025 Aug 27]. Available from: https://science.nasa.gov/mission/hubble/science/science-highlights/mapping-the-cosmic-web/
10. Encyclopaedia Britannica. Supermassive black hole. Britannica [Internet]. [cited 2025 Aug 27]. Available from: https://www.britannica.com/science/supermassive-black-hole
11. Collins GW. The Virial Theorem in Stellar Astrophysics. Tucson (AZ): Pachart Press; 1978. ISBN: 978-0-912918-13-6.
12. NJIT Physics. Virial equilibrium lecture slide [Internet]. [cited 2025 Aug 27]. Available from: https://slideplayer.com/slide/14853326/90/images/4/Virial+Theorem+and+Collapse.jpg
13. Diemer B, et al. An analysis of halo density profiles. arXiv [Preprint]. 2012 Dec 12 [cited 2025 Aug 27]. Available from: https://arxiv.org/abs/1212.2980
14. Villaescusa-Navarro F, et al. CAMELS: Cosmology and Astrophysics with Machine Learning Simulations. ReadTheDocs [Internet]. [cited 2025 Aug 27]. Available from: https://camels.readthedocs.io/en/latest/organization.html?highlight=profiles+illustristng+l25n256

15. Harwood M. Density of a vacuum. The Physics Factbook [Internet]. 1999 [cited 2025 Aug 27]. Available from: https://hypertextbook.com/facts/1999/MesahHarwood.shtml

16. Rubin Observatory. Vera C. Rubin Observatory [Internet]. [cited 2025 Aug 27]. Available from: https://rubinobservatory.org/about/vera-rubin