Describing the Galaxy-Halo Relationship Using Machine Learning

Gabrielle-Marie Cosmiano, University of Pittsburgh

Advisors: Dr. Rachel Mandelbaum and Yesukhei Jagvaral

Artificial Intelligence and Physics Summer Undergraduate Research Program

Carnegie Mellon University, Pittsburgh, PA

Introduction

★ In modern cosmological models, there is much more dark matter than ordinary matter. Therefore, the suggested structure of the Universe is composed of Dark Matter Halos, Subhalos, and Galaxies. Additionally, since there is much more dark matter compared to baryonic matter, we can indirectly make observations on dark matter which can in turn, help us answer questions about galaxies and their properties. This structure of the Universe suggests that there is a relationship (galaxy-halo relationship) between dark matter halos and the galaxies formed within them.

★ For this project, we chose to study the galaxy-halo relationship by using a machine learning technique called **regression** to model the relationship between the **stellar half-mass radius** and various masses such as stellar mass, mass of dark matter, gas mass, and total subhalo mass using data from the **Illustris-TNG simulation**.

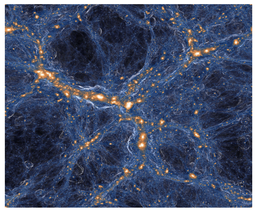


Fig. 1: *Image of cosmic web from the Illustris-TNG simulation. The orange in the picture represents collapsed dark matter structures and the blue represents shock surfaces.*

Methodology

★ **The Illustris-TNG simulation ("The Next Generation")** —An improved hydrodynamical simulation that allows us to test our predictions of galaxy properties and compare these predictions to observational data.

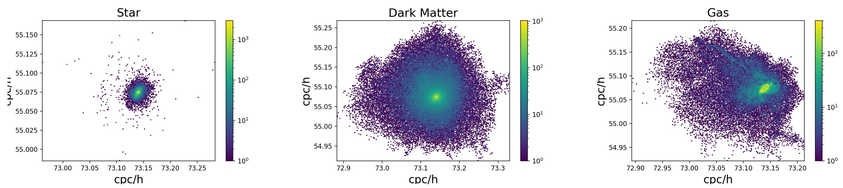


Fig. 2: *Here we are using data from the group catalogues in the Illustris-TNG Simulation to read in a single galaxy and the coordinates of all its constituent particles . We can see that there is a larger abundance of dark matter (DM) and gas particles compared to stellar particles in this galaxy and that DM and gas particles span out a wider physical area compared to stellar particles.*

★ **Multilayer Perceptron (MLP)**— A type of feedforward artificial neural networks.

★ **Regression**— Process for modeling a relationship between input features and an outcome.

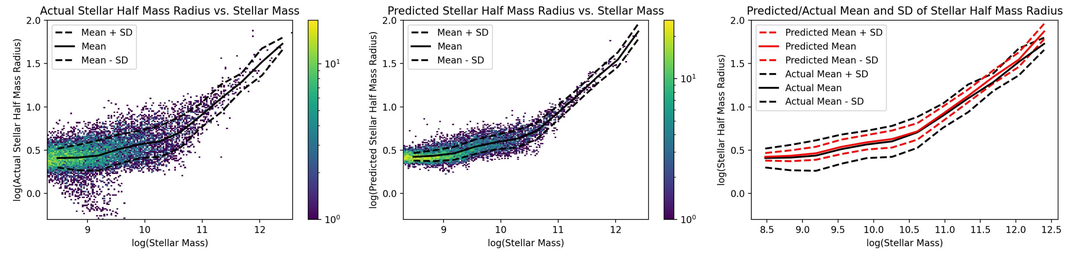
★ **Stellar Half-Mass Radius**— A measure of the physical size of the galaxy.

Methodology

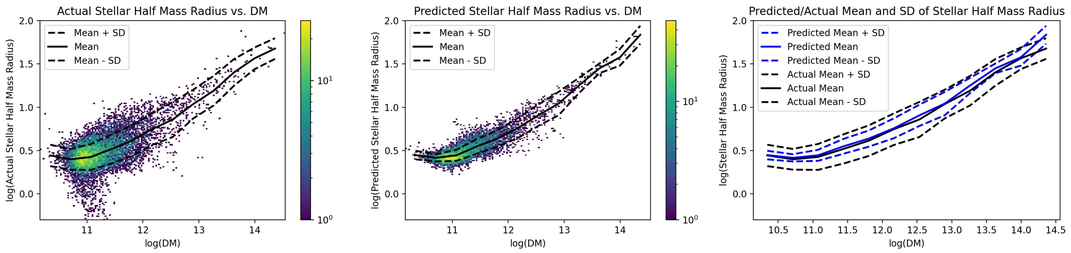
★Using data from group catalogues in the Illustris-TNG simulation, we split our data into 50% training and 50% testing subsets. We fed the various mass features into the model in order to produce a predicted output of the stellar half mass radius property.

★We then calculated the mean and the standard deviation for both the actual and predicted stellar half mass radius and compared the results.

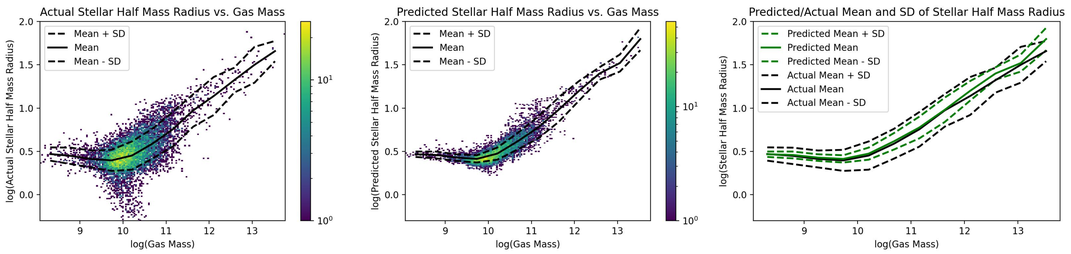
Results



*First row shows predicted/actual stellar half mass radius vs. **stellar mass**—third plot compares the mean and mean±SD for both predicted and actual stellar half mass radius with the input feature of stellar mass.*



*Second row shows predicted/actual stellar half mass radius vs. **the mass of dark matter**—third plot compares the mean and mean±SD for both predicted and actual stellar half mass radius with the input feature of DM.*



*Third row shows predicted/actual stellar half mass radius vs. **gas mass**—third plot compares the mean and mean±SD for both predicted and actual stellar half mass radius with the input feature of gas mass.*

***Note*:** The plots shown here are a representative subset of the results. Plots for the input feature of the total subhalo mass is not shown.

Results

★Predicted mean ± SD is much narrower than the actual mean ± SD, so this could indicate that the simulation does not capture the scatter very well. This is due to the simplicity of the **MLP** (model predicts mean correctly which minimizes the scatter).

★Actual mean and predicted mean produced by the model are close—this could indicate that the model is learning the relationship between the various masses and the stellar half mass radius.

Future Endeavors

★Implement other galaxy properties to strengthen stellar half mass radius prediction.

★Implement and use other machine learning techniques to address the shortcomings in the predicted scatter (optimize log-likelihood).

Acknowledgements

A special thanks to Dr. Mandelbaum and Yesukhei for guiding me in this project and for giving me the resources needed to learn and grow as a researcher. Thank you to the NSF AI Institute for funding this research project.

References

1. Brownlee, Jason. "Train-Test Split for Evaluating Machine Learning Algorithms." *Machine Learning Mastery*, 26 Aug. 2020, https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/.
2. de Santi, Natalí SM, et al. "Mimicking the halo–galaxy connection using machine learning." *Monthly Notices of the Royal Astronomical Society* (2022).
3. Dylan Nelson for the TNG Collaboration. "The Illustristng Project." *IllustrisTNG - Main*, https://www.tng-project.org/.
4. Nelson, Dylan, et al. "First results from the IllustrisTNG simulations: the galaxy colour bimodality." *Monthly Notices of the Royal Astronomical Society* 475.1 (2018): 624-647.

Carnegie Mellon University