

Estimating Dark Matter Halo Masses using Graph Neural Networks in Simulated Galaxy Clusters

NIKHIL GARUDA 

¹*Steward Observatory, University of Arizona, 933 N Cherry Ave, Tucson, AZ, 85721-0009, USA*

ABSTRACT

Predicting dark matter halos is a critical aspect of understanding the large-scale structure of the universe. This study focuses on predicting dark matter halo masses using stellar masses within galaxy clusters, leveraging the TNG-Cluster simulation from the IllustrisTNG as the foundational dataset. The primary objective is to develop a model that can accurately determine halo masses based on stellar mass and clustering/positional information using Graph Neural Networks (GNNs). Unlike traditional machine learning models like random forests, our GNN captures the information-rich substructure of galaxy clusters by using spatial and kinematic relationships between galaxy neighbors. We showed that our GNN model predicts dark matter halo masses more accurately (RMSE=0.273) than classical machine learning models. Future work will extend this approach across different simulations and real observational datasets to further validate the GNN model's ability to generalize.

1. INTRODUCTION

In the Λ CDM cosmological model (Peebles 1984; Bullock & Boylan-Kolchin 2017), Galaxies form and co-evolve in tandem with their dark matter halos over cosmic timescales. Although this galaxy–halo connection has been characterized using detailed cosmological hydrodynamic simulations, many important trends can be described via simple relationships and models (Wechsler & Tinker 2018). For example, a galaxy's stellar mass M_* scales with its dark matter halo mass M_{halo} , and this stellar-to-halo mass relation (SHMR) depends on whether the galaxy resides in a central or satellite subhalo.

While M_* is observable, M_{halo} must often be inferred indirectly via the SHMR due to observational constraints. For example, galaxy clusters—the most massive gravitationally bound objects in the Universe—are dark matter dominated, but their total mass must be measured via gravitational lensing (Clowe et al. 2006; Vegetti et al. 2024), the Sunyaev-Zel'dovich effect (Birkinshaw 1999; Marriage et al. 2011; Bleem et al. 2015), and/or visible wavelength proxies (e.g., galaxy richness, intracluster light, etc; Rykoff et al. (2014); Sampaio-Santos et al. (2021)). However, these methods are unable to fully leverage galaxy substructure within clusters to estimate their dark matter halo masses.

Therefore, we present a graph neural network (GNN) algorithm (Scarselli et al. 2009) for predicting M_{halo} for galaxies in simulated cluster environments. Compared to primitive machine learning (ML) methods like random forests (Agarwal et al. 2018), a GNN can learn

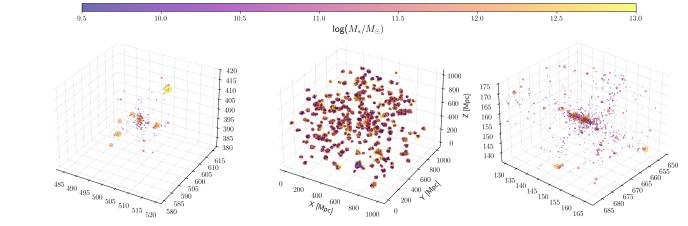


Figure 1. Spatial distribution of halos within the TNG-Cluster simulation. The middle panel shows the full simulation, and the left and right panels highlight two example galaxy clusters. The boundaries of these clusters are marked as blue and red boxes in the middle panel.

the substructure in neighboring galaxies and thereby improve halo mass predictions. Our results using the GNN demonstrate significant performance gains on the training, validation, and an independent test set.

2. DATA

The simulation data we use are large-volume, cosmological, gravo-magnetohydrodynamical simulations from the IllustrisTNG simulation suite (Nelson et al. 2019). We specifically use the TNG-Cluster (Nelson et al. 2024) simulation, a collection of zoom-in simulations centered 352 of the most massive subhalos (i.e., galaxy clusters). Our dataset is based on the SUBFIND (Springel et al. 2001) subhalo catalogs that were obtained from snapshot 99 ($z = 0$), focusing on the high-resolution components of the zoom-in simulation. We adopt cosmological parameters from (Planck Collaboration et al. 2016), using $H_0 = 67.74 \text{ km s}^{-1} \text{ Mpc}^{-1}$ for consistency with the IllustrisTNG simulation suite. The distribution of sub-

halos along with two example clusters in TNG-Cluster is shown in Figure 1.

To mimic astronomical observations of galaxies, we project the galaxy clusters along the z axis, which is chosen to be the line of sight. This procedure bridges the gap between simulation data and spectroscopic observations, which typically capture two spatial dimensions (x, y) and line-of-sight velocities (v_z). We also apply quality cuts to the simulation in Table 1 to ensure a complete sample of massive, well-resolved galaxies. The selection thresholds for halo masses and stellar masses are chosen to ensure simulation completeness, minimizing the impact of poorly resolved galaxies.

The decision to use a 2D projection of the galaxy clusters is driven by our aim to mimic realistic observational conditions achievable with telescopes, which typically capture two spatial dimensions and line-of-sight velocities. This approach helps bridge the gap between simulation data and observational astronomy, making our analysis more applicable to real-world scenarios.

We split the TNG-Cluster data into training and validation sets by implementing a k -fold cross-validation strategy based on cluster IDs rather than traditional random splits. This method isolates subhalos according to their cluster IDs while ensuring that all subhalos from a single cluster remain within the same fold. One potential caveat of this method is that we do not include the contaminating structure along the line-of-sight from other clusters which might be in a different k -fold.

We test our ML models on an independent data set, the Illustris TNG300-1 hydrodynamic simulation (hereafter TNG300; Nelson et al. (2019)). The TNG-Cluster and TNG300 simulations use the same physics and have comparable resolutions (in the former’s zoom-in regions), but the two simulations are otherwise independent. When reporting the both the TNG-Cluster cross-validation results and TNG300 test set results, we only consider galaxies within 10 Mpc of all clusters with $M_{\text{halo}} > 10^{14} M_{\odot}$ ($R_{200} \approx 1$ Mpc).

3. METHODS/EXPERIMENTS

The primary objective of our study is to estimate M_{halo} from M_* . Building on the work of (Larson et al. 2024), we train ML models on galaxies and dark matter halos from TNG-Cluster to probe cluster environments. We discuss performance metrics in §3.1. We present baseline models to establish a benchmark in §3.2, and then use a more sophisticated GNN architecture in §3.3.2. This comparison allows us to evaluate the added value of galaxy environments using GNNs to predict M_{halo} .

3.1. Evaluation Metrics and Loss Function

To quantify the performance of our predictive models, we utilize several metrics that will be shown in Table 2. Our simple models are trained to minimize the Mean Squared Error (MSE), while the GNN is optimized against the Gaussian negative log-likelihood (via the “Momentum Networks” approach that combines an MSE and log-variance term; Jeffrey & Wandelt (2020)). The loss is based on two terms: on the logarithmic stellar mass $\|\hat{y} - y\|^2$, and the squared difference between the predicted and measured variance $\|\hat{\Sigma} - (\hat{y} - y)^2\|^2$.

The validation and test set performance is quantified using the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), linear correlation coefficient (R^2), Normalized Median Absolute Deviation (NMAD), Bias, and Outlier Fraction (f_{outlier}).

3.2. Baseline Models

To establish a benchmark for subsequent comparisons with our GNN model, we use Random Forest (RF) regression (Ho 1995) as a baseline model due to its capability to handle complex non-linear relationships between features. To further augment the simple RF model, we compute an overdensity parameter (Δ_G), defined as the sum of stellar masses within a specified radius R_{max} . The RF models are configured with 100 estimators using scikit-learn (Pedregosa et al. 2011), one of which utilizes M_* , and one which uses both M_* and Δ_G as features.

We also run the same experiments done for RF using XGBoost. But one thing to note about XGBoost models is that they are more prone to overfitting. Although these baseline models are not expected to outperform the GNN, they serve as a valuable reference for measuring the improvements from more complex, structure-aware techniques.

3.3. Graph Neural Networks

A graph is a mathematical structure comprising a set of objects and the relationships between those objects. For this paper, we construct a cosmic graph to represent all subhalos and their pair-wise relationships.

3.3.1. Graph Features

M_* are represented as nodes, and pairs of subhalos are connected with edges if they are separated by less than L , the linking length equal to 3 Mpc (Wu et al. 2024) to predict M_{halo} . We do not permit nodes to connect to themselves (i.e., no self-loops). The ordering of subhalos is irrelevant, as graphs are invariant (and their nodes are equivariant) to permutations of the node indices. We will typically denote nodes or node properties using a single index, such as i , and edges or edge properties with two indices, such as ij .

Table 1. Summary of cuts applied to the TNG-Cluster data. Here, N_* refers to the number of stellar particles, M_\odot refers to solar mass, R_{200} refers to the virial radius of the subhalo.

Sample	Number of Subhalos
Full TNG-Cluster catalog	10,378,451
— within mass cuts - $N_* > 50$; $\log(M_*/M_\odot) > 9.5$; $\log(M_{\text{halo}}/M_\odot) > 10.5$	154,120
— within $< 10 \times R_{200}$ of the cluster halo	127,165
Selection Criteria - $\log(M_{\text{halo}}/M_\odot) > 11$; within 10 Mpc of the cluster halo	
TNG-Cluster cross-validation	60,756
TNG300 Test Set	34,689

We provide two edge features to incorporate both the spatial and kinematic separations of galaxies: the squared Euclidean distance between pairs of galaxy positions ($\|\bar{x}^{(i)} - \bar{x}^{(j)}\|^2$), and pairs of relative line-of-sight velocities ($\|\bar{v}^{(i)} - \bar{v}^{(j)}\|^2$).

3.3.2. Graph Architecture/Model

The core architecture of our GNN leverages message passing (Battaglia et al. 2016, 2018) between neighboring nodes, dynamically updating both node and edge states during the learning process. This allows the network to incorporate data from every pair of neighboring nodes and the attributes of their connected edge features. A pooling layer finally aggregates the data from these interactions returning to each node, allowing for simultaneous predictions of each node’s halo mass in the batch. This is based on the architecture described in Villanueva-Domingo et al. (2023); Wu et al. (2024).

Our GNN uses 8 unshared layers and 3 sequential layers based on the architecture provided in Figure 2. Each GNN layer comprises a two-layer MLP with 16 hidden channels, SiLU activation functions (Hendrycks & Gimpel 2016), and 16 outputs. These networks operate over the edges connected to each node, and a max pooling layer is used to aggregate information from the edges back onto the node, effectively capturing the neighboring features. The output from each node is then concatenated with its initial feature (stellar mass) which then is passed through a single 3-layer MLP. Finally, the GNN predicts two quantities (Jeffrey & Wandelt 2020): M_{halo} , and the expectation value of the M_{halo} log variance at a given M_* .

We employ the METIS algorithm to partition the training set into 48 parts. This batching procedure allows us to manage the computation demands for the large graph dataset. This partitioning, implemented through the `ClusterLoader` class in PyTorch Geometric (Chiang et al. 2019; Fey & Lenssen 2019), significantly reduces memory usage while preserving the local neighborhood structure necessary for effective GNN training. We train the model using the AdamW opti-

mizer (Loshchilov & Hutter 2019) with an initial learning rate of 10^{-2} and a weight decay of 10^{-4} for regularization. We use a learning rate scheduler that reduces the rate by a factor of 0.2 if the validation loss does not improve by 10^{-3} within 15 consecutive epochs. An early stopping mechanism halts training if no improvement in validation loss is observed within 35 epochs. We set the maximum number of training epochs to 300. It takes 20 minutes to run this model on a Nvidia A6000 GPU with inference time less than 1 second.

4. RESULTS

Table 2 compares model performance for predicting M_{halo} from galaxies residing in clusters for the validation and test datasets. We additionally show the scatter of M_{halo} in the first row, which represents the most naive “prediction” of the sample mean. Below, we present the results for the baseline models and GNN models. In Figure 3, we display scatter plots of the true versus predicted masses for the TNG-Cluster cross-validation data set (top) and TNG300 test set (bottom).

The simplest RF model exhibits high error and very low predictive power. In fact, for the TNG300 test set, the simplest RF model produces an even higher error than the scatter inherent to the data. We ascribe this to the RF model’s significant negative bias (i.e., systematic underprediction). When we augment the RF model with Δ_G , the performance improves, demonstrating that galaxy environments contain vital information for the SHMR. Nonetheless, the RF models systematically underpredict M_{halo} for the highest-mass galaxies and yield high error.

GNNs greatly outperform RF models, as indicated by the right-most panels of Figures 3. Running the same experiments using XGBoost, we find a significant improvement over RF but not enough to surpass GNNs. We find that the GNN performance on the training and validation sets translates to accurate predictions on the independent test set. For nearly all metrics in Table 2, the GNN outperforms the RF models for cross-validation and test sets.

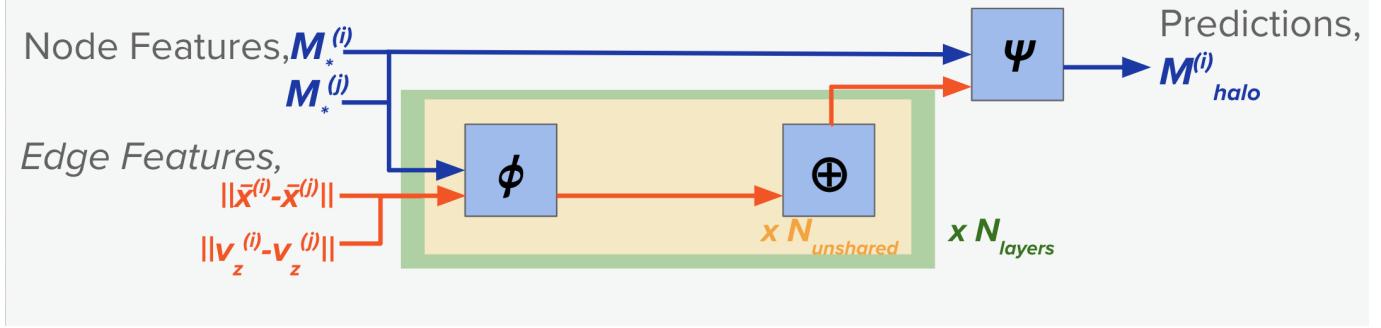


Figure 2. Flow diagram of GNN architecture used for halo mass prediction. The GNN processes node features (x_i, x_j) and edge features (ϵ_{ij}) through multiple unshared layers, where each layer applies learnable functions, ϕ , which are implemented as MLPs. These unshared layers operate in parallel across the graph structure. A pooling layer then aggregates (\oplus) the information from these interactions back into each node. Subsequent repetitions of these GNN layers can give it more representational power. Finally, the output MLP, ψ , combines node features and aggregated edge features to predict each node’s halo mass.

Table 2. Validation and test set performance for all models. The best metrics are underlined.

Model	RMSE	MAE	R^2	Bias	$f_{outlier}$	NMAD
TNG-Cluster cross-validation						
(Always predict mean)	0.542	0.396	0	0	0.019	0.479
RF: M_*	0.489 ± 0.002	0.382 ± 0.003	0.186 ± 0.011	-0.067 ± 0.006	0.008 ± 0.001	0.463 ± 0.005
RF: $M_* + \Delta_G$	0.385 ± 0.002	0.301 ± 0.002	0.490 ± 0.007	-0.124 ± 0.004	0.008 ± 0.000	0.367 ± 0.002
GNN	<u>0.273 ± 0.010</u>	<u>0.209 ± 0.009</u>	<u>0.745 ± 0.019</u>	<u>-0.085 ± 0.027</u>	<u>0.013 ± 0.002</u>	<u>0.246 ± 0.013</u>
TNG-300 test set						
(Always predict mean)	0.466	0.351	0	0	0.021	0.422
RF	0.468 ± 0.009	0.365 ± 0.014	0.199 ± 0.030	-0.200 ± 0.017	<u>0.009 ± 0.003</u>	0.456 ± 0.033
RF: $M_* + \Delta_G$	0.344 ± 0.003	0.256 ± 0.003	0.567 ± 0.007	-0.048 ± 0.001	0.022 ± 0.001	0.293 ± 0.006
GNN	<u>0.242 ± 0.013</u>	<u>0.184 ± 0.010</u>	<u>0.785 ± 0.023</u>	<u>-0.039 ± 0.034</u>	<u>0.014 ± 0.002</u>	<u>0.217 ± 0.014</u>

Note: The intrinsic scatter in M_{halo} ranges from 0.42 (at $\log(M_{halo}) = 11 M_\odot$) to 0.33 (at $\log(M_{halo}) = 13 M_\odot$) dex in TNG-Cluster and 0.48 dex to 0.19 dex in TNG300 respectively.

5. DISSCUSION

5.1. Comparison against previous work

For a long time, there have been previous studies that use dark matter halo properties to obtain central and halo baryonic properties (Agarwal et al. 2018). There have been feature importance studies that determined which feature is the most important using reinforcement learning to predict M_* from M_{halo} with deep neural networks (Moster et al. 2021). Kamdar et al. (2016) used semi-analytical models to qualitatively analyze the extent of the influence of dark matter halo properties on galaxies but this has been done only with smaller simulations. CNNs and GNNs have also been used to do galaxy cluster mass estimation using M_{halo} (Yan et al. 2020; Wu & Jespersen 2023).

Larson et al. (2024) is the first study that explicitly predicts M_{halo} from M_* and morphological parameters using GNN+CNN. This is the first time we are predict-

ing M_{halo} from M_* for high mass regions - galaxy clusters using GNNs.

5.2. Model performance as a function of local environment

In Figure 4, we show the RMSE as a function of distance from the cluster center for the GNN and RF (M_* and Δ_G) models; the GNN significantly outperforms the RF across all distance bins. Notably, the RF model performance suffers for galaxies closer to the center of the cluster. One potential explanation for this discrepancy is that the RF does not account for the dense cluster environment, where interactions such as tidal stripping can lead to a significant loss of M_{halo} . Due to line-of-sight effects, not all galaxies at small projected distances experience significant tidal stripping. In contrast, the GNN model outperforms the RF by leveraging information from galaxy pairwise distances and line-of-sight velocities.

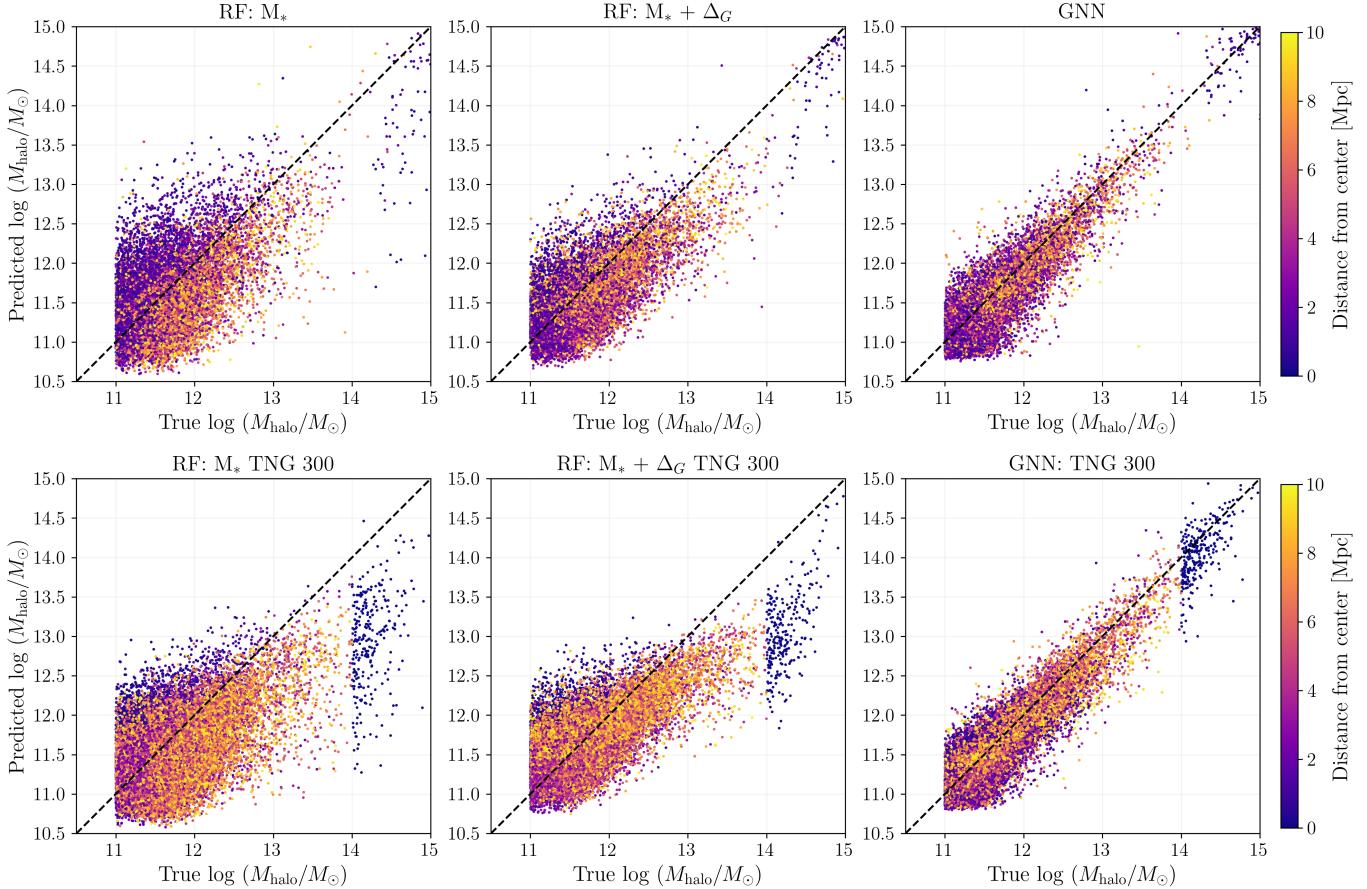


Figure 3. Predicted versus true M_{halo} for all the models. The top panels present TNG-Cluster validation set results for a single k -fold, while the bottom panels present TNG300 test set results. From left to right panels, we show the random forest (RF) performance using only M_* as a feature, the RF using M_* and overdensity parameter (Δ_G), and the GNN using M_* and graph connectivity.

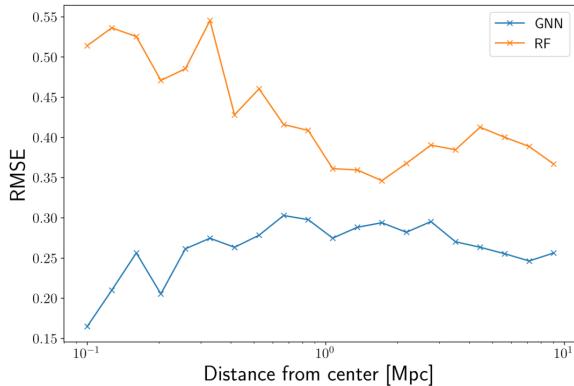


Figure 4. Validation set RMSE as a function of distance from the cluster center. Results are shown for the GNN (blue) and RF with M_* and Δ_G (orange).

5.3. Domain shift and other limitations

ML models often exhibit domain shift (Sun et al. 2016; Kouw & Loog 2019), i.e., they perform poorly when applied to different datasets than the ones they are trained

on. In our work, the model performance is comparable between the TNG-Cluster cross-validation and TNG300 test datasets. Because the training and independent test datasets use the same underlying physics from the IllustrisTNG simulation suite, they may be robust to domain shift. The GNN model exhibits particularly strong performance on the test set, which may demonstrate that it has learned a symbolic relationship that is generalizable across domains (e.g., Cranmer et al. (2020)). Future extensions to our work can include tests on cosmological simulations that have different physical models and different initial conditions (e.g., Schaye et al. (2023)).

We may also apply our to real observational datasets that could also pose its own sets of challenges and limitations. At present, the models are currently not applicable to real observations. We would need to train on different redshifts and account for corrections that could account for how the SEDs change with redshift. We will also need to account for realistic observational effects, including contamination from bad z_{phot} or line-of-sight contamination due to $2d$ projection, sources missing

from photometric catalogs, and aperture measurements. Another limitation is also the fact that we trained using zoom-in simulation.

6. CONCLUSIONS AND FUTURE WORK

We predict M_{halo} for simulated galaxies characterized by their stellar masses and 2D projected positions and line-of-sight velocities (i.e., x, y, v_z) using the TNG-Cluster simulation as a training dataset and TNG300 as a test set. This work studies simple Random Forest (RF) models and more sophisticated graph neural networks (GNNs). We report the following conclusions:

1. The GNN model significantly outperforms the RF model, even when the latter is provided galaxy overdensity as a parameter. These results are shown in Figure 3 and Table 2. Thus, GNNs capture the underlying spatial relationships and substructure within clusters, which traditional models fail to do.
2. The GNN maintains its predictive power when tested on the independent TNG300 dataset, demonstrating that the model generalizes across the IllustrisTNG simulation suite.

Despite our promising results, models trained on one simulation may face challenges when applied to other simulations or real observational data. Future work will need to account for observational effects like contaminating galaxies in projection, missing data, and photometric redshift uncertainties, as well as broader concerns about domain shift in ML (see e.g. Ćiprijanović et al. (2021)). Testing the model on simulations with different physical models or initial conditions can further validate the applicability of ML algorithms. Aside from additional validation on other cosmological simulations, we will test on observational data using published M_{halo} estimates for well-known galaxy clusters (e.g., Lotz et al. (2017); Weaver et al. (2024)). With upcoming telescopes like the Roman Space Telescope (Spergel et al. 2015) and Rubin Observatory (Collaboration 2012), we will be able to study GNN applications to large galaxy cluster samples in the wide-field domain.

REFERENCES

- Agarwal, S., Davé, R., & Bassett, B. A. 2018, Monthly Notices of the Royal Astronomical Society, 478, 3410, doi: [10.1093/mnras/sty1169](https://doi.org/10.1093/mnras/sty1169)
- Battaglia, P. W., Pascanu, R., Lai, M., Rezende, D., & Kavukcuoglu, K. 2016, Interaction Networks for Learning about Objects, Relations and Physics, arXiv, doi: [10.48550/arXiv.1612.00222](https://doi.org/10.48550/arXiv.1612.00222)
- Battaglia, P. W., Hamrick, J. B., Bapst, V., et al. 2018, Relational Inductive Biases, Deep Learning, and Graph Networks, arXiv, doi: [10.48550/arXiv.1806.01261](https://doi.org/10.48550/arXiv.1806.01261)
- Birkinshaw, M. 1999, Physics Reports, 310, 97, doi: [10.1016/S0370-1573\(98\)00080-5](https://doi.org/10.1016/S0370-1573(98)00080-5)
- Bleem, L. E., Stalder, B., de Haan, T., et al. 2015, The Astrophysical Journal Supplement Series, 216, 27, doi: [10.1088/0067-0049/216/2/27](https://doi.org/10.1088/0067-0049/216/2/27)
- Bullock, J. S., & Boylan-Kolchin, M. 2017, Annual Review of Astronomy and Astrophysics, 55, 343, doi: [10.1146/annurev-astro-091916-055313](https://doi.org/10.1146/annurev-astro-091916-055313)
- Chiang, W.-L., Liu, X., Si, S., et al. 2019, in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 257–266, doi: [10.1145/3292500.3330925](https://doi.org/10.1145/3292500.3330925)
- Ćiprijanović, A., Kafkes, D., Downey, K., et al. 2021, Monthly Notices of the Royal Astronomical Society, 506, 677, doi: [10.1093/mnras/stab1677](https://doi.org/10.1093/mnras/stab1677)
- Clowe, D., Bradač, M., Gonzalez, A. H., et al. 2006, The Astrophysical Journal, 648, L109, doi: [10.1086/508162](https://doi.org/10.1086/508162)
- Collaboration, L. D. E. S. 2012, Large Synoptic Survey Telescope: Dark Energy Science Collaboration, arXiv, <https://arxiv.org/abs/1211.0310>
- Cranmer, M., Sanchez-Gonzalez, A., Battaglia, P., et al. 2020, Discovering Symbolic Models from Deep Learning with Inductive Biases, arXiv, doi: [10.48550/ARXIV.2006.11287](https://doi.org/10.48550/ARXIV.2006.11287)
- Fey, M., & Lenssen, J. E. 2019, Fast Graph Representation Learning with PyTorch Geometric, arXiv, doi: [10.48550/ARXIV.1903.02428](https://doi.org/10.48550/ARXIV.1903.02428)
- Hendrycks, D., & Gimpel, K. 2016, Gaussian Error Linear Units (GELUs), arXiv, doi: [10.48550/ARXIV.1606.08415](https://doi.org/10.48550/ARXIV.1606.08415)
- Ho, T. K. 1995, in Proceedings of 3rd International Conference on Document Analysis and Recognition, Vol. 1, 278–282 vol.1, doi: [10.1109/ICDAR.1995.598994](https://doi.org/10.1109/ICDAR.1995.598994)
- Jeffrey, N., & Wandelt, B. D. 2020, Solving High-Dimensional Parameter Inference: Marginal Posterior Densities & Moment Networks, arXiv, doi: [10.48550/ARXIV.2011.05991](https://doi.org/10.48550/ARXIV.2011.05991)

- Kamdar, H. M., Turk, M. J., & Brunner, R. J. 2016, Monthly Notices of the Royal Astronomical Society, 455, 642, doi: [10.1093/mnras/stv2310](https://doi.org/10.1093/mnras/stv2310)
- Kouw, W. M., & Loog, M. 2019, An Introduction to Domain Adaptation and Transfer Learning, arXiv. <https://arxiv.org/abs/1812.11806>
- Larson, A. J., Wu, J. F., & Jones, C. 2024, Predicting Dark Matter Halo Masses from Simulated Galaxy Images and Environments, arXiv. <https://arxiv.org/abs/2407.13735>
- Loshchilov, I., & Hutter, F. 2019, Decoupled Weight Decay Regularization, arXiv, doi: [10.48550/arXiv.1711.05101](https://doi.org/10.48550/arXiv.1711.05101)
- Lotz, J. M., Koekemoer, A., Coe, D., et al. 2017, The Astrophysical Journal, 837, 97, doi: [10.3847/1534-4357/837/1/97](https://doi.org/10.3847/1534-4357/837/1/97)
- Marriage, T. A., Acquaviva, V., Ade, P. A. R., et al. 2011, The Astrophysical Journal, 737, 61, doi: [10.1088/0004-637X/737/2/61](https://doi.org/10.1088/0004-637X/737/2/61)
- Moster, B. P., Naab, T., Lindström, M., & O’Leary, J. A. 2021, Monthly Notices of the Royal Astronomical Society, 507, 2115, doi: [10.1093/mnras/stab1449](https://doi.org/10.1093/mnras/stab1449)
- Nelson, D., Pillepich, A., Ayromlou, M., et al. 2024, Astronomy & Astrophysics, 686, A157, doi: [10.1051/0004-6361/202348608](https://doi.org/10.1051/0004-6361/202348608)
- Nelson, D., Springel, V., Pillepich, A., et al. 2019, Computational Astrophysics and Cosmology, 6, 2, doi: [10.1186/s40668-019-0028-x](https://doi.org/10.1186/s40668-019-0028-x)
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825
- Peebles, P. J. E. 1984, The Astrophysical Journal, 284, 439, doi: [10.1086/162425](https://doi.org/10.1086/162425)
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, Astronomy & Astrophysics, 594, A13, doi: [10.1051/0004-6361/201525830](https://doi.org/10.1051/0004-6361/201525830)
- Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, The Astrophysical Journal, 785, 104, doi: [10.1088/0004-637X/785/2/104](https://doi.org/10.1088/0004-637X/785/2/104)
- Sampaio-Santos, H., Zhang, Y., Ogando, R. L. C., et al. 2021, Monthly Notices of the Royal Astronomical Society, 501, 1300, doi: [10.1093/mnras/staa3680](https://doi.org/10.1093/mnras/staa3680)
- Scarselli, F., Gori, M., Ah Chung Tsoi, Hagenbuchner, M., & Monfardini, G. 2009, IEEE Transactions on Neural Networks, 20, 61, doi: [10.1109/TNN.2008.2005605](https://doi.org/10.1109/TNN.2008.2005605)
- Schaye, J., Kugel, R., Schaller, M., et al. 2023, Monthly Notices of the Royal Astronomical Society, 526, 4978, doi: [10.1093/mnras/stad2419](https://doi.org/10.1093/mnras/stad2419)
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report. <https://arxiv.org/abs/1503.03757>
- Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, Monthly Notices of the Royal Astronomical Society, 328, 726, doi: [10.1046/j.1365-8711.2001.04912.x](https://doi.org/10.1046/j.1365-8711.2001.04912.x)
- Sun, B., Feng, J., & Saenko, K. 2016, in Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI’16 (Phoenix, Arizona: AAAI Press), 2058–2065
- Vegetti, S., Birrer, S., Despali, G., et al. 2024, Space Science Reviews, 220, 58, doi: [10.1007/s11214-024-01087-w](https://doi.org/10.1007/s11214-024-01087-w)
- Villanueva-Domingo, P., Villaescusa-Navarro, F., Anglés-Alcázar, D., et al. 2023, Inferring Halo Masses with Graph Neural Networks, doi: [10.3847/1538-4357/ac7aa3](https://doi.org/10.3847/1538-4357/ac7aa3)
- Weaver, J. R., Cutler, S. E., Pan, R., et al. 2024, The Astrophysical Journal Supplement Series, 270, 7, doi: [10.3847/1538-4365/ad07e0](https://doi.org/10.3847/1538-4365/ad07e0)
- Wechsler, R. H., & Tinker, J. L. 2018, Annual Review of Astronomy and Astrophysics, 56, 435, doi: [10.1146/annurev-astro-081817-051756](https://doi.org/10.1146/annurev-astro-081817-051756)
- Wu, J. F., & Jespersen, C. K. 2023, Learning the Galaxy-Environment Connection with Graph Neural Networks, arXiv. <https://arxiv.org/abs/2306.12327>
- Wu, J. F., Jespersen, C. K., & Wechsler, R. H. 2024, How the Galaxy-Halo Connection Depends on Large-Scale Environment, arXiv. <https://arxiv.org/abs/2402.07995>
- Yan, Z., Mead, A. J., Van Waerbeke, L., Hinshaw, G., & McCarthy, I. G. 2020, Monthly Notices of the Royal Astronomical Society, 499, 3445, doi: [10.1093/mnras/staa3030](https://doi.org/10.1093/mnras/staa3030)