Linking Warm Dark Matter to Merger Tree Histories via Deep Learning Networks

ILEM LEISHER [1,2,3] PAUL TORREY [4,5,6] ALEX M. GARCIA [4,5,6] JONAH C. ROSE [7,8]

FRANCISCO VILLAESCUSA-NAVARRO [7,8] ZACHARY LUBBERTS [9] ARYA FARAHI [10,11] STEPHANIE O'NEIL [12,8]

XUEJIAN SHEN [3] OLIVIA MOSTOW [4] NITYA KALLIVAYALIL [4,5] DHRUV ZIMMERMAN [13] DESIKA NARAYANAN [13,14]

AND MARK VOGELSBERGER [2,3,15,16]

[1] Grinnell College, 1115 8th Ave, Grinnell, IA 50112, USA

[2] Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3] Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[4] Department of Astronomy, University of Virginia, 530 McCormick Road, Charlottesville, VA 22904 USA

[5] The NSF-Simons AI Institute for Cosmic Origins, USA

[6] Virginia Institute for Theoretical Astronomy, University of Virginia, Charlottesville, VA 22904, USA

[7] Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA

[8] Department of Physics, Princeton University, Princeton, NJ 08544, USA

[9] Department of Statistics, University of Virginia, Charlottesville, VA

[10] Departments of Statistics and Data Science, University of Texas at Austin, Austin, TX 78757, USA

[11] The NSF-Simons AI Institute for Cosmic Origins, University of Texas at Austin, Austin, TX 78712, USA

[12] Department of Physics & Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA

[13] Department of Astronomy, University of Florida, 211 Bryant Space Sciences Center, Gainesville, FL 32611 USA

[14] Cosmic Dawn Center at the Niels Bohr Institute, University of Copenhagen and DTU-Space, Technical University of Denmark

[15] The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Cambridge, MA 02139, USA

[16] Fachbereich Physik, Philipps Universität Marburg, D-35032 Marburg, Germany

ABSTRACT

Dark matter (DM) halos form hierarchically in the Universe through a series of merger events. Cosmological simulations can represent this series of mergers as a graph-like "tree" structure. Previous work has shown these merger trees are sensitive to cosmology simulation parameters, but as DM structures, the outstanding question of their sensitivity to DM models remains unanswered. In this work, we investigate the feasibility of deep learning methods trained on merger trees to infer Warm Dark Matter (WDM) particles masses from the DREAMS simulation suite. We organize the merger trees from 1,024 zoom-in simulations into graphs with nodes representing the compete merger history of galaxies within a $z = 0$ root node merger history, and edges denoting hereditary links. We vary the complexity of the node features that are included in the graphs ranging from a single node feature up through an array of several galactic properties (e.g., halo mass, star formation rate, etc.). We train a Graph Neural Network (GNN) to predict the WDM mass using the graph representation of the merger tree as input. We find that the GNN can predict the mass of the WDM particle with varying success (R^2 from 0.07 to 0.95), depending on the nodes, edges, and node features used in the graphs. We extend the same methods to astrophysical parameters A_{SN1}, A_{SN2}, and A_{AGN} regarding supernovae and active galactic nuclei feedback, finding success in inferring the supernovae parameters. The GNN can even infer the WDM mass from merger tree histories without any node features, indicating that the structure of merger trees alone inherits information about the cosmological parameters of the simulations from which they form.

Corresponding author: Ilem Leisher

ilemleisher@gmail.com

1. INTRODUCTION

Within the standard Λ cold dark matter (ΛCDM) model of cosmology, dark matter (DM) is assumed to be cold and collisionless. As a whole, CDM functions exceptionally well for the large-scale structures in the Universe (Spergel et al. 2007; de Blok et al. 2008; Dubinski & Carlberg 1991; Dalal & Kochanek 2002; Van Waerbeke, L. et al. 2005), but creates tensions at the sub-galactic level (Nadler et al. 2021b; Moore 1994; Boylan-Kolchin et al. 2011; Papastergis, E. et al. 2015; Oman et al. 2015; Klypin et al. 1999; Moore et al. 1999). These tensions can be alleviated via modified treatments of either the galaxy formation model, or modifications to the cosmological framework itself.

Due to the difficult nature of observing DM and its growth over time, cosmological simulations are extremely valuable to investigations of DM models. In particular, the recent use of Machine Learning (ML) for analyzing data and making inferences within these simulations has proven to be a powerful tool for the study of DM (Fluri et al. 2021; Villaescusa-Navarro et al. 2022a). For example, the CAMELS project (Villaescusa-Navarro et al. 2021a, 2023, 2022b) and the DREAMS project (Rose et al. 2025) both employ ML methods in simulation suites tracking DM particles to study structure evolution as a function of cosmology and astrophysics.

The traditional ML tool for this task has been Convolutional Neural Networks (CNNs), which have been successfully used to infer cosmological parameters (Rose et al. 2023; Villaescusa-Navarro et al. 2021b,c). However, the capabilities of CNNs are limited by their inputs, which must occupy a linear, Euclidean space. This forces input data into a grid structure, overlooking non-linear spatial dependencies the data may have, and potentially resulting in a worse performance of the model. This has led to investigations of other ML methods, such as Graph Neural Networks (GNNs; Scarselli et al. 2009), which offer architectures to capture complex non-linear spatial dependencies. GNNs learn from input data occupying graph spaces, which preserve such non-linear spatial relationships in the data. They have been shown to be an effective model for inferring both cosmological (Villanueva-Domingo & Villaescusa-Navarro 2022) and astrophysical information (Jespersen et al. 2022; Garuda et al. 2024; Chuang et al. 2024), in some cases more so than CNNs (Jespersen et al. 2022). Learning from non-linear spatial dependencies allow GNNs to extract more information from certain datasets than CNNs can.

There have been several examples of GNNs demonstrating their efficacy in inferring astrophysical and cosmological information within simulations (Villanueva-Domingo & Villaescusa-Navarro 2022; Garuda et al. 2024; Beck et al. 2019; Shao et al. 2023; de Santi et al. 2023a,b; Massara et al. 2023; Jespersen et al. 2022; Chuang et al. 2024). Such capabilities have provided insight into how the evolution of particular structures (e.g., galaxy distributions, merger trees) are influenced by simulation parameters. For the study of DM, the examination of merger trees is particularly interesting, as these are structures that trace a DM halo's growth through merger events. Though recent work has successfully used GNNs to infer both cosmological parameters and galaxy properties from merger trees (Jespersen et al. 2022; Chuang et al. 2024; Huang et al. 2025), it has so far not been shown if a GNN can infer DM properties (such as the particle mass) from merger trees. In this paper, we test this possibility by training a GNN on merger tree histories from cosmological simulations.

In this work we use a Warm Dark Matter (WDM) simulation suite from the DREAMS project (Rose et al. 2025) to provide us with cosmological and astrophysical parameters for our GNN. A WDM model is a simple category of DM models, in that it has only one parameter, particle mass (Bode et al. 2001), making it particularly compatible with ML methods (Rose et al. 2023; Lin et al. 2024) and thus fitting for our purposes. In contrast to CDM, a WDM model assumes somewhat lighter DM particles leading to higher streaming velocities in the early universe that wash out density fluctuations on small scales. Previous work has assessed whether WDM can accurately describe the DM features we observe in the Universe or not (Wang & White 2007; Lovell et al. 2012; Viel et al. 2013; Lovell et al. 2014; Fitts et al. 2019). Primarily, the WDM particle mass has been constrained via comparisons to the observed satellite luminosity function in the Local Group (Kennedy et al. 2014; Murgia et al. 2017; Nadler et al. 2021a), with the Lyman-alpha forest from high redshift quasars (Viel et al. 2013; Iršič et al. 2017), and between small halo properties in strongly lensed systems (Ritondale et al. 2019; Gilman et al. 2019). The strongest constraints have been placed on the particle mass using a combination of methods: Nadler et al. (2021b) and Enzi et al. (2021) found upper and lower limits of 9.7 keV and 6.0 keV, respectively.

Although the model potentially alleviates CDM tensions such as the production of cored density profiles (Bode et al. 2001), these constraints are incompatible with observational properties of dwarf galaxies, which require a WDM particle mass of ~ 2 keV (Lovell et al. 2014; Fitts et al. 2019). Thus, WDM is an unlikely candidate for the entire DM population in the Universe (Enzi et al. 2021; Nadler et al. 2021b). Regardless, demonstrating that a GNN can infer WDM properties from merger tree histories paves the way for doing so with more realistic models in the future, while also creating a framework where generative models for merger trees can appropriately reflect the cosmology from which they are drawn. Following the work of Lin et al. (2024), we also infer three astrophysics parameters within the WDM suite that control supernovae (SN) and active galactic nuclei (AGN) feedback. This further tests the capabilities of our model as well as the information contained within the merger trees.

The structure of this paper is outlined as follows. In Section 2, we describe the simulation suite from which we obtain our merger trees and our chosen GNN model. In Section 3, we present our results from training the GNNs on a variety of galaxy properties to infer the WDM and astrophysics parameters, focusing on the WDM parameter as it has the most interesting results. In Section 4, we discuss our results and investigate WDM sensitivities contained in the merger trees. Finally, in Section 5, we state our conclusions.

2. METHODS

Our goal is to build a framework to perform graph-based inference using GNNs that take the galaxy merger trees as input in order to predict cosmological and simulation parameters. This requires (i) a simulation dataset with varied astrophysical and cosmological parameters, (ii) a procedure for converting merger trees into graph objects, and (iii) an appropriate GNN architecture. We describe each of these elements in this section.

2.1. *Simulations*

Our data is from the WDM Milky Way zoom-in suite within the DaRk mattEr and Astrophysics with Machine learning and Simulations (DREAMS) project (Rose et al. 2025)[1]. This suite is well-suited for our study as it has a broad range of WDM, cosmology, and astrophysics parameter variations while also maintaining high baryon mass resolution, which allows for reasonably detailed resolution of galaxy merger trees (sub)halo

---

---

formation histories. The suite contains 1,024 hydrodynamical simulation zoom-in boxes run using the IllustrisTNG galaxy formation model (Pillepich et al. 2017; Weinberger et al. 2018), which itself is based on the Illustris galaxy formation model (Vogelsberger et al. 2013; Torrey et al. 2014), and advanced with the moving-mesh code AREPO (Springel 2010; Springel et al. 2019; Weinberger et al. 2020). Initial conditions are generated using MUSIC (Hahn & Abel 2011), which evolves a uniform particle distribution through time using second-order Lagrangian perturbation equations. Each simulation spans $z = 127$ to $z = 0$ and includes 91 snapshots from $z = 15$ to $z = 0$. A (single) merger tree is created from each simulation using SUBLINK (Rodriguez-Gomez et al. 2015) from subhalos identified with SUBFIND (Springel et al. 2001). The mass of subhalos in the suite ranges from 7×10^{11} $M_\odot$ to 2.5×10^{12} $M_\odot$. Compared to the TNG50-1 simulation, the highest resolution box of the IllustrisTNG suite (Nelson et al. 2019; Pillepich et al. 2019), the DREAMS zoom-in suites have $\sim 5\times$ as many MW-mass galaxies at comparable baryon mass resolution (a factor of ~ 2 lower).

2.1.1. *Parameters*

Although the DREAMS suites are primarily based on the IllustrisTNG physics model, there are a number of systematic parameter variations of both the DM physics and astrophysics of the simulations. In this section, we detail the parameter variations within the WDM suite.

The simulations in this suite have fixed cosmological parameters that are consistent with Planck Collaboration et al. (2016): $\Omega_b = 0.046, h = 0.691, \Omega_m = 0.302, \Omega_\Lambda = 0.698, \sigma_8 = 0.839$. In the suite we use, there are three varied astrophysical parameters, A_{SN1}, A_{SN2}, A_{AGN}, and one varied cosmological parameter, M_{WDM}. We now briefly explain each varied parameter (we refer the reader to Rose et al. 2025 Section 2.1 for a complete description of the model).

The A_{SN1} and A_{SN2} parameters are dimensionless scaling factors in the IllustrisTNG SN feedback model. This model contains two main components, both of which depend on these scaling factors. The first component is the mass-loading factor η_w, which depends on the second component, stellar wind velocity v_w, such that $\eta_w \propto e_w / v_w^2$. e_w is the specific energy available for generating winds, which scales based on A_{SN1}, according to the equation

$$e_w = A_{SN1} \bar{e}_w \left[f_{w,Z} + \frac{1 - f_{w,Z}}{1 + (Z/Z_{w,\mathrm{ref}})^{\gamma_{w,z}}} \right] \times N_{SNII} \; E_{SNII,51}. \quad (1)$$

In this equation, Z is the metallicity of star-forming gas, $f_{w,Z}$ is included to reduce the available energy when the metallicity is above $Z_{w,\mathrm{ref}}$, $\bar{e}_w$ is a dimensionless scaling factor set to 3.6 in the fiducial TNG model, N_{SNII} is the number of SN Type II per formed stellar mass, $E_{\mathrm{SNII},51}$ is the available energy per core collapse SNe, and A_{SN1} is our varied scaling factor parameter.

The second component, stellar wind velocity, is scaled by A_{SN2}, according to the equation

$$v_w = \max\left[A_{\mathrm{SN2}}\ \kappa_w\ \sigma_{\mathrm{DM}}\left(\frac{H_0}{H(z)}\right)^{1/3}, v_{w,\mathrm{min}}\right], \quad (2)$$

where $H(z)$ is the Hubble parameter, $v_{w,\mathrm{min}}$ is the minimum wind speed, κ_w is a dimensionless scaling (set to 7.4 in the fiducial model), σ_{DM} is the local one-dimensional DM velocity dispersion, and A_{SN2} is our varied scaling factor parameter.

The A_{AGN} parameter scales the fraction of energy transferred to nearby gases due to black hole accretion ($\epsilon_{f,\mathrm{high}}$). This is incorporated into the IllustrisTNG AGN feedback model via the feedback energy released from accreting black holes. Notably, black holes are not seeded in DM halos below $\sim 7 \times 10^7 M_\odot$ (Weinberger et al. 2018). The equation for this feedback energy is

$$\Delta\dot{E} = A_{\mathrm{AGN}}\epsilon_{f,\mathrm{high}}\epsilon_r\dot{\mathrm{M}}_{\mathrm{BH}}c^2, \quad (3)$$

where ϵ_r is the radiative efficiency, which is the canonical 0.1–0.2 of the accreted rest-mass energy that is released in the accretion process, $\dot{\mathrm{M}}_{\mathrm{BH}}$ is the black hole accretion rate, and A_{AGN} is our varied parameter.

The M_{WDM} parameter is the WDM particle mass. WDM is implemented into the simulations following a suppression of the matter power spectrum from Bode et al. (2001):

$$P_{\mathrm{WDM}}(k) = \beta(k)P_{\mathrm{CDM}}(k), \quad (4)$$

where $P_{\mathrm{WDM}}(k)$ and $P_{\mathrm{CDM}}(k)$ are the WDM and CDM power spectra, respectively. $\beta(k)$ is given by

$$\beta(k) = ((1 + (\alpha k)^{2.4})^{-5.0/1.2})^2, \quad (5)$$

and

$$\alpha = 0.048\ h^{-1}\ (M_{\mathrm{WDM}})^{-1.15}$$
$$\times \left(\frac{\Omega_{\mathrm{m}} - \Omega_{\mathrm{b}}}{0.4}\right)^{0.15}\left(\frac{h}{0.65}\right)^{1.3} \quad (6)$$

where Ω_{m} is the total matter density (fixed to 0.302 in this suite) and Ω_{b} is the total baryon density. The suppression factor models the free-streaming behavior of WDM particles, erasing small-scale power below a characteristic scale that is dependent on the assumed WDM particle mass.

These four parameters are varied following a Sobol sequence (Sobol' 1967) in order to uniformly sample across the parameter space. Sobol sampling provides dense coverage of the specified parameter space while improving convergence rate compared to purely random sampling. The WDM suite uses the following parameter sampling ranges:

$$0.033\ \mathrm{keV}^{-1} < 1/M_{\mathrm{WDM}} < 0.555\ \mathrm{keV}^{-1} \quad (7)$$
$$0.25 < \log_{10}(A_{\mathrm{SN1}}) < 4 \quad (8)$$
$$0.5 < \log_{10}(A_{\mathrm{SN2}}) < 2 \quad (9)$$
$$0.25 < \log_{10}(A_{\mathrm{AGN}}) < 4 \quad (10)$$

2.1.2. SubLink *Merger Trees*

The SubLink merger trees are constructed by assigning a descendant to each subhalo in a given box based on a three-step process detailed in Rodriguez-Gomez et al. (2015) and outlined here. First, for a given subhalo, all subhalos in the next snapshot backward in time that share common particles with the subhalo are tagged as candidates. Next, each candidate is given a score as a function of the binding energy rank of the common particles between the two particle-sharing subhalos. Finally, the candidate with the highest score is defined as the unique descendant of the given subhalo.

Each subhalo in SubLink trees is given pointers to five key subhalos (Springel et al. 2005). The first is the *first progenitor*: the progenitor of a given subhalo with the most massive history. The second is the *next progenitor*: the subhalo which shares the same descendant as a given subhalo, which has the next largest mass history. The third is the *descendant*: the unique subhalo that both has common particles with a given subhalo and has the least binding energy rank summed across these particles. The fourth is the *first subhalo in Friend-of-Friends group*: the main subhalo (defined as the one with the most massive history) from the Friend-of-Friends (FoF; Davis et al. 1985) group of a given subhalo. The final key subhalo is the *next subhalo in FoF group*: the next subhalo from the same FoF group, in order of decreasing mass history.

We walk each merger tree using a top-down approach moving backward in time, beginning with the MW subhalo at snapshot 90. We find this subhalo's first progenitor and all of its next progenitors, and then walk down the subtrees stemming from each of these subhalos, using a recursive function to repeat this action for every subhalo in the tree. Unlike previous work that has used trees with certain subhalos excluded (see Section 4.4 for

details; Jespersen et al. 2022; Chuang et al. 2024), we elect to include all subhalos in our trees. Throughout this process, we select data properties from both the available SubLink fields and the FoF group catalog associated with each subhalo, and save them in a manner that preserves the structure of the merger tree within a storage file.

2.2. Graph Neural Network

Graph representation of the merger trees preserves the galaxy assembly history information in a form that can be naturally ingested into GNNs. In this section, we describe the specific form of our merger tree-based graphs, the architecture and application of our GNN model, as well as the loss function we use.

2.2.1. Graphs

We convert SubLink merger trees into graph objects using the PyTorch Geometric package (Fey & Lenssen 2019), visualized in Figure 1. In this package, graph objects $G = (X, (I, E))$ are comprised of a node feature matrix X containing node features x, and a sparse adjacency tuple (I, E) containing edges directed backwards in time with edge indices encoded in I and optional edge features for each edge in E. We map each data property in a given storage file to a node feature in the corresponding graph, and we create pairs of edge indices by matching each subhalo with its unique descendant to satisfy the edge tuple. We do not include any edge features, as these represent information contained within the edges between two nodes (i.e., spatial distance between two nodes), whereas we are only interested in the existence of an edge between two nodes. Though this choice nominally translates to a loss of information, we believe it is fitting for our methods of investigating merger trees. For the cosmological and astrophysical parameter values M_{WDM}, A_{SN1}, A_{SN2}, A_{AGN} in each box, we normalize each value with $(A - \min(A))/(\max(A) - \min(A))$, where A is a parameter and $\min(A)$ and $\max(A)$ are the bounds of the corresponding sampling space. We attach these normalized values to each associated graph.

2.2.2. Architecture

We use a GNN model similar to CosmoGraph-Net[2] (Villanueva-Domingo 2022) used in Villanueva-Domingo & Villaescusa-Navarro (2022), which follows the Graph Network framework from Battaglia et al. (2018). Villanueva-Domingo & Villaescusa-Navarro
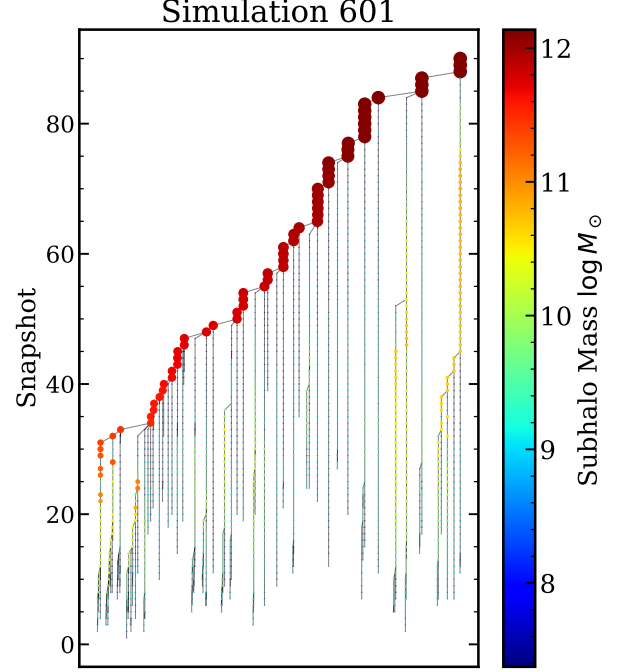
Figure 1. The graph representation of a merger tree from an example simulation within the DREAMS WDM MW zoom-in suite. The merger tree represents the merger history of a MW-mass subhalo beginning at $z = 0$ (snapshot 90) and ending at $z = 15$ (snapshot 0). All subhalos in the tree are included. The colors of the nodes, as well as the sizes of the nodes, represent the subhalo mass (ranging from 7×10^{11} M$_\odot$ to 2.5×10^{12} M$_\odot$).
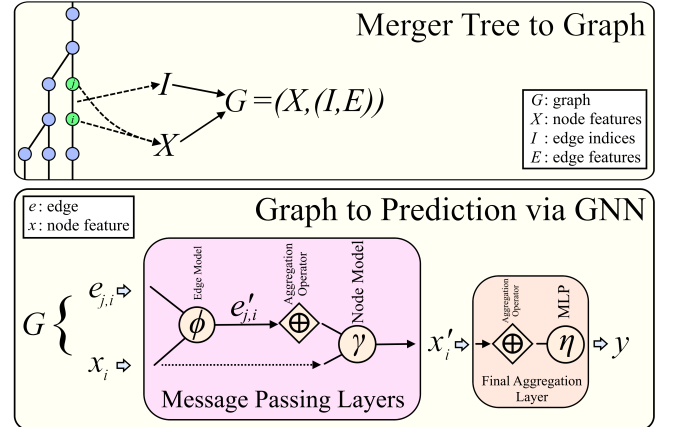
Figure 2. A schematic diagram of the tree-to-graph conversion and the GNN architecture. Node features and edges (consisting of edge indices and meaningless edge features) are encoded into a graph, which then pass through several message passing layers and a final aggregation layer, producing an output prediction.

(2022) use their GNN to both predict the power spectrum of a set of galaxies and infer the value of cosmological parameters. Thus, we choose to adopt the same GNN architecture with slight adaptations to facilitate compatibility of our merger tree graphs. We now briefly explain the model architecture as it applies to our analysis. For a more in-depth description of the architecture, we direct the reader to Villanueva-Domingo & Villaescusa-Navarro (2022), Section 2.3.

The GNN consists of an encoder graph block, hidden message-passing layers, and a final aggregation layer. The encoder graph block and message-passing layers use a neighborhood aggregation scheme (see Fey & Lenssen 2019, Section 2 for more details), incorporating edge and node models. Since we do not have edge features, the edge model exists only to satisfy the input requirements for the node model. In these layers, an edge model takes in the edge and node features and outputs new edge features, and a node model takes in the node and updated edge features and outputs new node features. Unlike Villanueva-Domingo & Villaescusa-Navarro (2022), we do not input any global features into either model, as we are concerned solely with the structural properties of the merger tree and the data features of the individual subhalos, neither of which are described with global properties. Thus, the node model produces updated node features x_i', such that:

$$x_i' = \gamma([x_i, \ \bigoplus_{j \in N_i} \phi(x_i, \ x_j, \ e_{j,i})]) \quad (11)$$

where x_i, x_j are two linked nodes, ϕ and γ are multilayer perceptrons (MLPs) which include a ReLU activation layer, $\phi(x_i, \ x_j, \ e_{j,i})$ represents the updated edge features, N_i is the neighborhood of node i (all the nodes with an edge connected to i), and $\bigoplus$ is a permutational invariant operator that aggregates all the information from neighboring nodes. Some examples of $\bigoplus$ are sum, mean, and maximum. We incorporate a multi-pooling layer in the node model to use all three possible aggregation operators before applying the MLP, which Villanueva-Domingo & Villaescusa-Navarro (2022) found produced slightly more accurate results.

The message passing layers also include a residual layer which adds the input of the layer to its output; a technique that has been shown to reinforce the loss function against chaotic behavior (Li et al. 2018). The final aggregation layer uses the same multi-pooling layer with its aggregation operator as the node model, an MLP, and a final linear layer, which outputs an array of predictions y for the parameter, such that:

$$y = \eta([\bigoplus_{i \in G} x_i']) \quad (12)$$

where η is an MLP, and $\bigoplus$ is defined as before. We show a diagram outlining this architecture as well as the graph construction in Figure 2.

2.2.3. Node features

Of the available data features associated with each subhalo, we create combinations of five node features to attach to our graphs:

1. **S**. The number of the snapshot in which the subhalo is found.

2. **M_DM**. Total mass of all DM particles that are bound to the subhalo.

3. **M_∗**. Total mass of all stellar components that are bound to the subhalo.

4. **M_G**. Total mass of all gas components that are bound to the subhalo.

5. **SFR**. The sum of the individual star formation rates of all gas cells in the subhalo.

We use these node features in specific to summarize both the baryonic (**M_G**, **M_∗**, **SFR**) and DM (**M_DM**) attributes of subhalos. **S** is required in all combinations of node features to maintain the temporal information of the tree structure. Although including data from more SUBFIND fields might produce more accurate inferences, we choose to maintain the focus of our work on testing the feasibility of our methods rather than optimizing quantitative results. For in-depth identification of WDM deterministic node features with GNNs, we direct the reader to Costanza et al. in preparation.

2.2.4. Loss Function

We use the same loss function used to perform likelihood-free inference of cosmological parameters by Villanueva-Domingo & Villaescusa-Navarro (2022). Our network is trained to predict the posterior mean and standard deviation of a parameter, which is done using a specific loss function given by Jeffrey & Wandelt (2020):

$$\begin{aligned} L \ = \ & \log \left(\sum_{i=1}^{N} \sum_{\alpha=1}^{A} \left(\theta_{\alpha,i} - \mu_{\alpha,i} \right)^2 \right) \\ & + \log \left(\sum_{i=1}^{N} \sum_{\alpha=1}^{A} \left[\left(\theta_{\alpha,i} - \mu_{\alpha,i} \right)^2 - \sigma_{\alpha,i}^2 \right]^2 \right), \quad (13) \end{aligned}$$

where N are batch elements, A are components, θ_a is the parameter, μ_a is the posterior mean, and σ_a is the standard deviation.

2.2.5. *Training*

We divide our merger tree dataset into 80% training, 10% validation, and 10% testing, using a seeded shuffle. We use a seeded shuffle because it means each unique GNN has identical dataset division, thus making comparisons between networks fair. Our GNN has four hyperparameters that we tune using OPTUNA (Akiba et al. 2019). Automated tuning is necessary given the sensitivity of GNN performance to, e.g., architecture depth and learning rates. OPTUNA does this by suggesting hyperparameters within a range for each trial that minimize the validation loss, meaning that each trial has a different set of values for these hyperparameters. The hyperparameters and their ranges are: learning rate (*lr*) 10^{-7} to 10^{-3}, weight decay (*wd*) 10^{-9} to 10^{-6}, number of layers (*nl*) 1 to 5, and hidden channels (*hc*) 2^n, with n from 6 to 9. We train each of our models for 50 trials of 1000 epochs of hyperparameter tuning for each unique combination of node features.

2.2.6. *Evaluation Metrics*

We use two metrics to evaluate the success of our model. The coefficient of determination is our primary metric, which is given by:

$$R^2 = 1 - \frac{\sum_i^N (y_{\text{true,i}} - y_{\text{pred,i}})^2}{\sum_i^N (y_{\text{true,i}} - \bar{y}_{\text{pred,i}})^2}, \quad (14)$$

where values closer to 1 mean more accurate results. We also calculate the root mean squared error (RMSE) as a secondary metric, defined such that:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_i^N (y_{\text{true,i}} - y_{\text{pred,i}})^2}, \quad (15)$$

with smaller values meaning more accurate results. In both equations, N is the number of trees in the testing set, $y_{\text{true,i}}$ is the true parameter value, $y_{\text{pred,i}}$ is the predicted parameter value, and $\bar{y}_{\text{pred,i}}$ is the average of the predicted parameter values. We include both R^2 and RMSE since they describe different metrics for the "accuracy" of the model (the correlation of variables and error, respectively). However, as both the R^2 and RMSE similarly and sufficiently evaluate the GNN's performance, we generally defer to showing just the R^2 in the text. For completeness, we show both metrics in figures and tables.

3. RESULTS

In this section, we present our results from training four GNN models to infer the cosmological parameter M_{WDM} and astrophysical parameters A_{SN1}, A_{SN2}, and A_{AGN}. We first train the models with just one node feature, and then with increasingly more node features. We focus on the results of the M_{WDM} model, as it performs the best and gives us the most interesting results. For extended results of the other three models, see Appendix A.

3.1. *Merger History as an Input to GNNs*

We begin testing our four GNN models by training them with just one node feature: snapshot number **S**. As the halos in merger trees evolve with time, having time dependence in the node features is essential. Including snapshot number accomplishes this, though notably, since we compute summary statistics across nodes, the snapshot number represents temporal information that is not necessarily the assembly of a halo with time. By training a model with the snapshot number alone, we are also implicitly testing whether the number and/or frequency of mergers in a given halo's history carries information about the underlying DM and astrophysics of our simulations. We explicitly test these sensitivities later in Section 4.1.

We train each model for 50 trials of 1000 epochs of hyperparameter tuning following the process in Section 2.2.5. Figure 3 shows the inferred parameter values compared against the true parameter values. Each plot shows the results from a different model (labeled in the bottom right of each panel). The dotted lines show the one-to-one relation: models that are closer to this relation are more accurately predicting the true values.

The top left panel of Figure 3 shows the results of the WDM GNN. We find that the GNN appears to successfully infer WDM particle mass for masses $\lesssim 4$ keV. The box in the top right of the plot shows a zoom-in of this low mass regime. Although there is some variance, the general trend in this regime roughly follows the diagonal, demonstrating its success. The success of the GNN in predicting WDM particle mass at $\lesssim 4$ keV is remarkable given that the input features of the graph is only the snapshot number. This suggests that the number and/or frequency of merging events that a halo undergoes in its assembly is sensitive to the underlying DM physics. Indeed, the sensitivity of the merger rates to DM physics makes qualitative sense, as many of the progenitors of the $z = 0$ halos are low mass systems (see, e.g., Figure 1). The growth of these low mass systems is being preferentially suppressed by the initial matter power spectrum variations. After ~ 8 keV, however, the model consistently predicts WDM masses below the actual values. We suspect that the model cannot do well in the $\gtrsim 8$ keV range due to constraints from the simulation resolution, which we discuss further in Section 4.2.
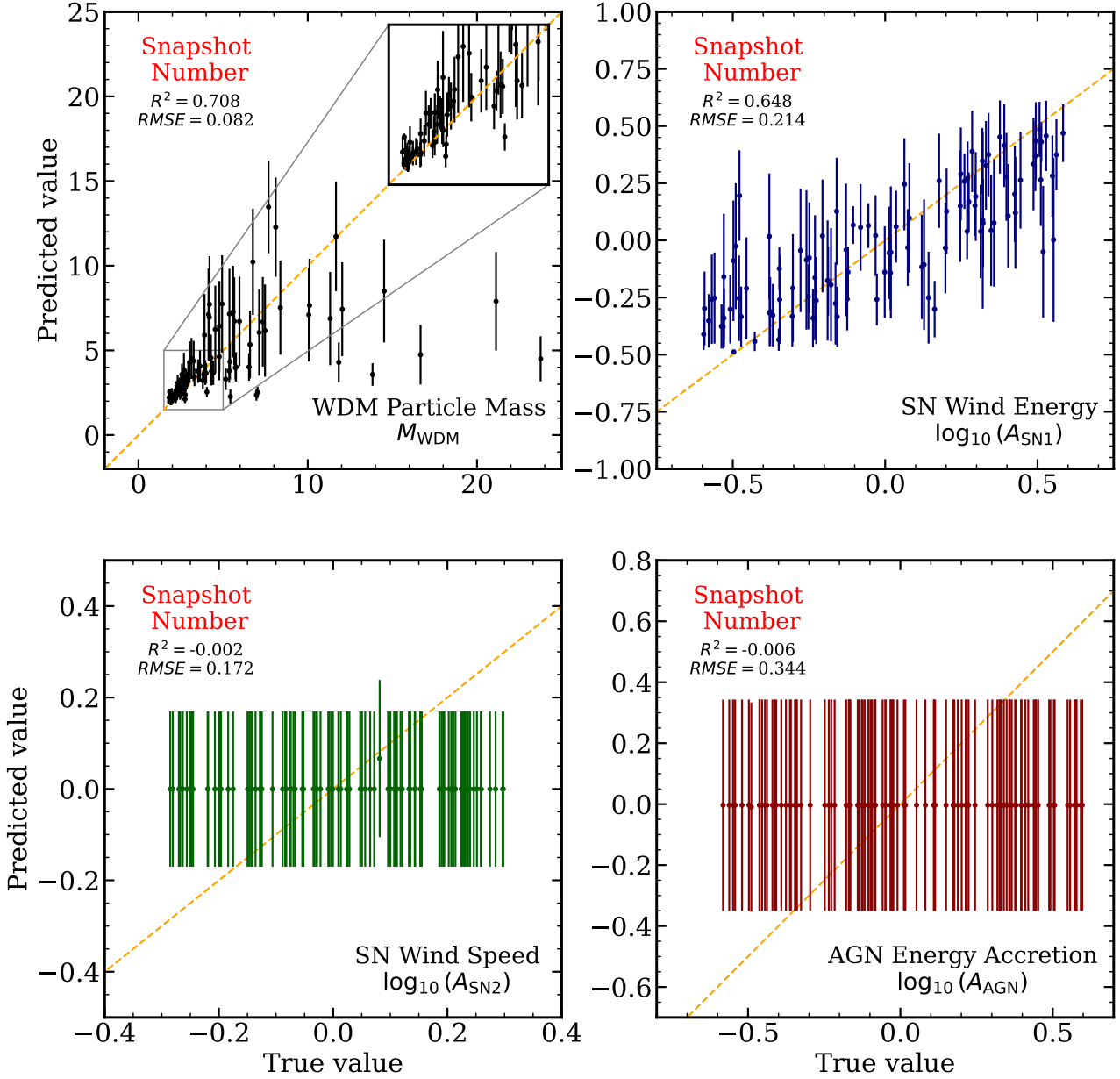
Figure 3. The plotted results of training the four GNN models on merger trees with one node feature attached: subhalo snapshot number. Each plot shows the inferred parameter values versus the true parameter values. The parameter inferred is labeled at the bottom right of each plot. The keys at the top left of each plot refer to the node feature that was used in training the model to produce the plot (see Section 2.2.3 for details). The dotted lines spanning the diagonals are where perfectly accurate predictions would fall. The inference of the WDM parameter is successful, particularly for lower masses, with a R^2 value of 0.708 and a RMSE of 0.082, as well as the SN1 parameter, with a R^2 value of 0.648 and a RMSE of 0.214. However, the A_{SN2} and A_{AGN} models both fail to make accurate predictions, as they only predict the mean value.

The top right panel of Figure 3 shows the results from inferring A_{SN1}. We find that the GNN infers the supernova wind energy parameter with some accuracy, making predictions that generally follow a diagonal trend, though with a relatively large degree of variance. Unlike the WDM inference, the model does well across the entire parameter range. The bottom left and bottom right panels show the results from inferring A_{SN2} and A_{AGN}, respectively. Both plots show the same behavior: the models predict the mean value repeatedly (with the exception of a single simulation in the A_{SN2} model). This demonstrates a complete failure of these models to infer their parameters from the information we have trained them with.

The lack of success in our model predicting the both the A_{SN2} and A_{AGN} parameters is perhaps unsurprising. In low mass halos and at early times, the impact of variations in the supernova winds (A_{SN2}) is curtailed by the minimum wind speed of the TNG model (see Equation 3). The lack of distinction in the wind speeds at high redshift and in low mass halos means that many of the DM halos in our trees should be entirely unimpacted by variations in A_{SN2}. Similarly, black holes are not seeded in DM halos below a halo mass of $\sim 7 \times 10^7 M_\odot$ (Weinberger et al. 2018). AGN feedback is therefore not present in a large number of halos in the trees. Then, once AGN feedback is present in the halos, the A_{AGN} variations play a self-regulatory effect on the total feedback budget from the black holes (see Appendix A of Garcia et al. In Preparation), making A_{AGN}'s impact on the DM halos weak. Therefore, we do not expect significant variations to the DM merger events by modifying either A_{SN2} or A_{AGN}, consistent with our results. However, as we explore in the next Section, A_{SN2} and A_{AGN} do become more important for the *baryonic* properties of the galaxies.

We quantify the performance of the models by calculating the coefficients of determination (see Section 2.2.5). The results from the M_{WDM} model produce a R^2 value of 0.708, A_{SN1} produces a R^2 value of 0.648, A_{SN2} produces a R^2 value of -0.002, and A_{AGN} produces a R^2 value of -0.006. These values support the findings above, showing that only the M_{WDM} and A_{SN1} models successfully learn to infer their parameters with some accuracy.

### 3.2. *Addition of Subhalo Features to GNNs*

The previous subsection indicates that the structure of the merger tree histories themselves (i.e. the trees labeled only with snapshot number) contains some information about the WDM particle mass as well as the supernova wind energy (A_{SN1}), but limited-to-no discriminatory information about the other astrophysical parameters. In this subsection, we investigate models trained with additional node features which reflect galaxy properties that are more directly influenced by the assumed astrophysics. We first examine the benefit of the addition of physical properties of the galaxy and host halo in the GNN's ability to predict the WDM particle mass. Then, using the same procedure, we present the best-performing models for the astrophysics parameter variations.

We train our WDM GNN model with four progressively richer combinations of the node features listed in Section 2.2.3, as follows: $(\mathbf{S}, \mathbf{M_{DM}})$, $(\mathbf{S}, \mathbf{M_{DM}}, \mathbf{M_*})$, $(\mathbf{S}, \mathbf{M_{DM}}, \mathbf{M_*}, \mathbf{M_G})$, $(\mathbf{S}, \mathbf{M_{DM}}, \mathbf{M_*}, \mathbf{M_G}, \mathbf{SFR})$. We choose the combinations such that DM properties ($\mathbf{M_{DM}}$) are added first, and then baryonic properties ($\mathbf{M_*}$, $\mathbf{M_G}$, $\mathbf{SFR}$). For each node combination, our network is trained for 50 trials of 1000 epochs of hyperparameter tuning.

Figure 4 shows the performance of the GNN trained on merger trees with the above node features. Each panel of Figure 4 shows the predicted WDM particle mass values against the actual WDM parameter values. Across all configurations, the models perform best in the low WDM particle mass regime ($\lesssim 4$ keV), where predictions closely track the true values with relatively low scatter. At larger WDM masses, the performance degrades for all networks with inferred values systematically undershooting the true values (consistent with the results of the previous section). At large WDM masses, all networks tend to predict WDM masses clustering around $\sim 10\text{-}15$ keV. This saturation effect is consistent with the resolution-driven limitations seen in earlier DREAMS analysis (expanded on in Section 4.2) and likely reflects the difficulty of distinguishing high-mass WDM models apart from one another where the suppressed portion of the matter power spectrum is not resolved (Rose et al. 2023).

Increasing the richness of the node features improves the network performance in several ways. The first panel demonstrates that adding $\mathbf{M_{DM}}$ dramatically improves the accuracy of the model (when compared with Figure 3), suggesting that including DM subhalo mass helps the model capture information about structure-suppression that is not encapsulated through the overall merger tree structure itself. In the second panel, we see that adding $\mathbf{M_*}$ to the $(\mathbf{S}, \mathbf{M_{DM}})$ configuration seems to produce a plot qualitatively similar to the previous, indicating it has little effect on the accuracy of the inferred values. In the third panel, we incorporate $\mathbf{M_G}$, which reduces the prevalence of over-predictions (i.e., above the diagonal) for masses $\gtrsim 10$ keV. Finally, in
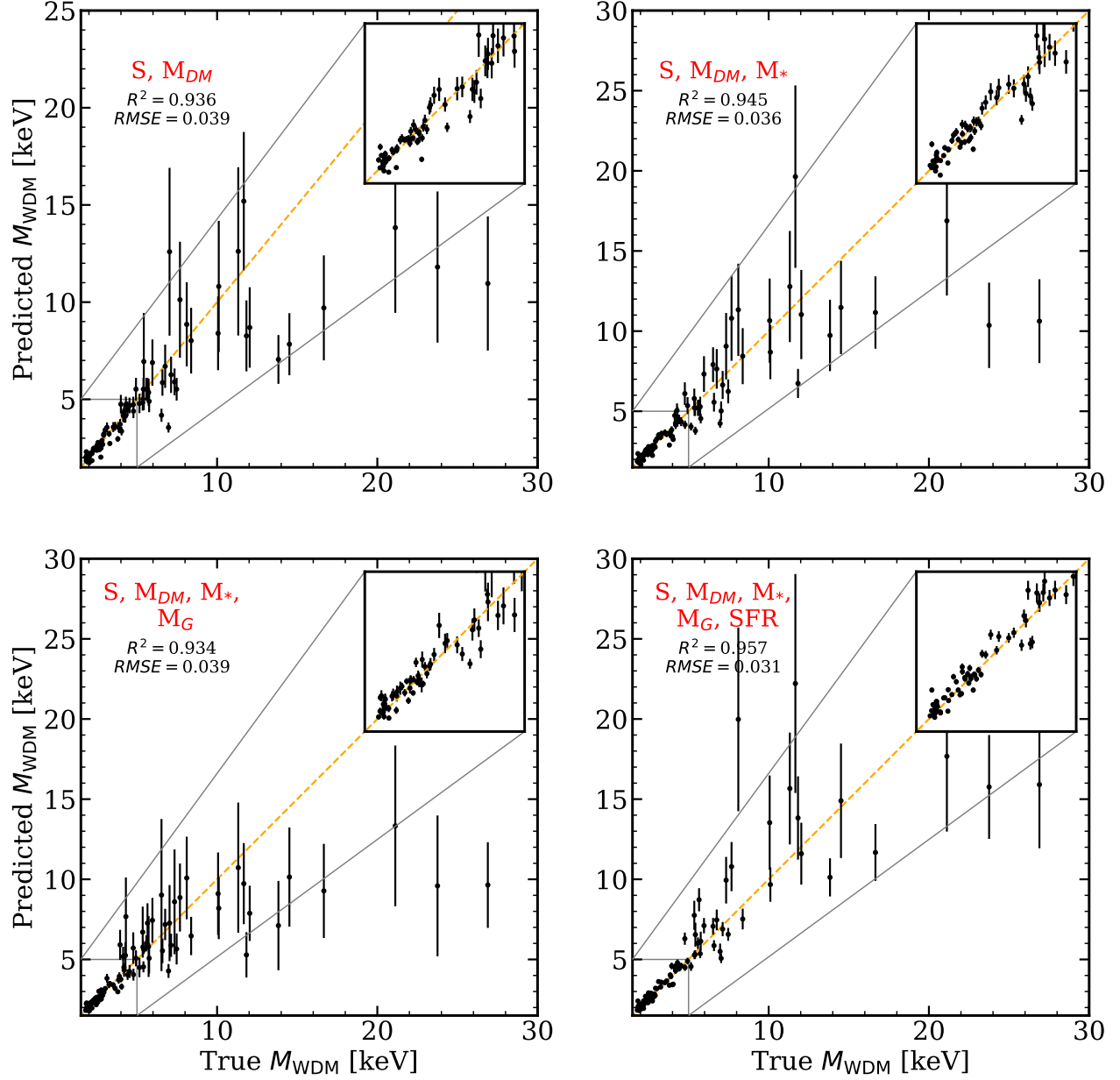
Figure 4. The plotted results of training the WDM GNN model on four different combinations of node features from the merger trees. Each plot shows the inferred WDM particle mass versus the true WDM particle mass. The keys at the top left of each plot refer to the node features that were used in training the model to produce the plot (see Section 2.2.3 for details). The dotted line spanning the diagonals are where perfectly accurate predictions would fall. The model successfully infers the WDM particle mass for all of the node combinations, performing roughly the same for each combination of node features. The best performance, in the bottom right, has a R^2 value of 0.957 and a RMSE of 0.031, demonstrating highly accurate predictions.

the fourth panel, we add **SFR**, which generally increases the predicted values for masses $\gtrsim 9$ keV, which brings some predictions closer to the diagonal but also causes some large over-predictions around ~ 9 keV. Overall, it does not significantly alter the qualitative results. These trends generally demonstrate that augmenting the node feature set with baryonic and DM halo properties progressively improves the GNN's ability to identify the employed WDM model, though the fundamental limitations associated with predicting WDM properties for the largest WDM mass scales remain.

We quantify the success of our network across each node combination by calculating the coefficients of determination (see Equation 14). We obtain coefficients of determination of 0.936 from (**S**, **M$_{\mathrm{DM}}$**), 0.945 from (**S**, **M$_{\mathrm{DM}}$**, **M$_*$**), 0.934 from (**S**, **M$_{\mathrm{DM}}$**, **M$_*$**, **M$_{\mathrm{G}}$**), and 0.957 from (**S**, **M$_{\mathrm{DM}}$**, **M$_*$**, **M$_{\mathrm{G}}$**, **SFR**). These results support our interpretation that the GNN is very successful across all four node combinations, though they also highlight a quantitative decrease in performance with the addition of **M$_{\mathrm{G}}$**. As this decrease is very slight, we attribute it to the GNN's variance from testing to testing, and not necessarily as an indication that this specific feature negatively influences the model. Overall, both the qualitative and quantitative metrics indicate that the GNN successfully learns to infer WDM particle mass with a high degree of accuracy.

Following the same procedure described above, we trained additional GNN models to infer the three astrophysical feedback parameters: A_{SN1}, A_{SN2}, and A_{AGN} (see Section 2.1.1 for definitions). For brevity, we present only the best-performing node feature combinations here, but the results of each node feature combination can be found in Appendix A. Each model underwent 50 trials of hyperparameter tuning over 1,000 epochs, using the same four node-feature combinations as in our WDM experiments: (**S**, **M$_{\mathrm{DM}}$**), (**S**, **M$_{\mathrm{DM}}$**, **M$_*$**), (**S**, **M$_{\mathrm{DM}}$**, **M$_*$**, **M$_{\mathrm{G}}$**), (**S**, **M$_{\mathrm{DM}}$**, **M$_*$**, **M$_{\mathrm{G}}$**, **SFR**). The best-performing results for each parameter are presented in Figure 5. We also show the hyperparameters used in each result in Table 1.

The networks show a mixed ability to recover these feedback parameters. For A_{SN1}, the GNN achieves strong predictive performance ($R^2 = 0.987$) for (**S**, **M$_{\mathrm{G}}$**, **M$_*$**, **M$_{\mathrm{G}}$**, **SFR**), indicating that galaxy properties such as stellar mass and star formation rate are sensitive to supernova wind energy injection. The A_{SN2} model performs slightly worse ($R^2 = 0.880$) for (**S**, **M$_{\mathrm{G}}$**, **M$_*$**, **M$_{\mathrm{DM}}$**, **SFR**), but still shows a clear correlation between predicted and true values. This suggests that signatures of the supernova wind velocity scaling are present, albeit with slightly increased variance compared to A_{SN1}. In

contrast, the model for A_{AGN} fails to learn meaningful relationships ($R^2 = 0.102$) with any node combination, with predictions not following any trend. This lack of predictive power is consistent with expectations: our Milky Way–mass zoom-in suite predominantly samples halos below the stellar-mass scale where AGN feedback becomes important (Weinberger et al. 2018), limiting the amount of A_{AGN} information encoded in the subhalo merger histories.

4. DISCUSSION

In this section, we conduct additional tests of merger tree structure and discuss the trends in our results from inferring the four parameters.

### 4.1. *Merger Tree Structure without Temporal Information*

The results of our WDM GNN trained on just one node feature (see Section 3.1) demonstrate that the network can infer WDM particle mass with snapshot number as the only attached information. The snapshot number represents the merger tree's time dependence, essentially assigning a time to each node feature. By including no other node features, the network is being trained on the structure of the merger tree histories alone which includes, for example, the total number of subhalos/mergers that contributed to the formation of the $z = 0$ halo. The success of the network with the absence of additional node features suggests that there is information contained in the structure of a subhalo's merger tree history regarding the WDM particle mass within the subhalo.

We further investigate the merger tree structural information by training our WDM GNN on merger trees without time dependence. We do this by creating graph objects with only a placeholder node feature $x = [1, 1, \ldots]$ attached to satisfy our GNN requirements. Without the addition of snapshot number as a node feature, we have stripped away the temporal element of the trees, thus training our GNN on the number of nodes in the graph and the associated edge connections. We train the model for 50 trials of 1000 epochs of hyperparameter tuning. The results are shown in Figure 6. We find that the model is able to infer the mass of the WDM particle, though with significantly less accuracy than previous results. It is particularly successful at low masses, but, unlike the results in Figure 3, there is also high uncertainty in this regime. Following the previously seen trend of the M_{WDM} model, we also see it undershoot most of the predictions after ~ 7 keV.

We calculate a R^2 value of 0.509 and a RMSE of 0.107. This is the least successful our WDM network has performed. Still, since we do not see a total failure of the

Best Performances						
Parameter	R^2	RMSE	lr	wd	nl	hc
M_{WDM}	0.957	0.031	$1.889 \cdot 10^{-5}$	$5.139 \cdot 10^{-7}$	4	128
A_{SN1}	0.987	0.041	$7.862 \cdot 10^{-6}$	$4.352 \cdot 10^{-9}$	5	64
A_{SN2}	0.880	0.059	$1.063 \cdot 10^{-7}$	$2.826 \cdot 10^{-9}$	4	256
A_{AGN}	0.102	0.325	$1.778 \cdot 10^{-74}$	$1.196 \cdot 10^{-7}$	5	128

Table 1. Metrics of performance and hyperparameters from the best trial of the best-performing node combination for each parameter. The hyperparameters are learning rate lr, weight decay wd, number of layers nl, and hidden channels hc.
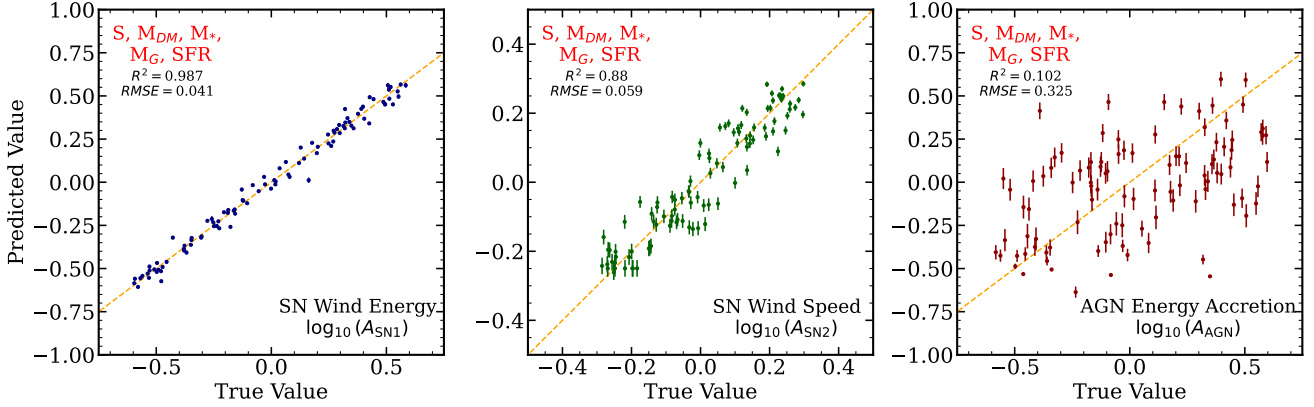
Figure 5. The plotted results of the best-performing node feature combinations from the A_{SN1}, A_{SN2}, and A_{AGN} GNN models. Each plot shows the inferred parameter value versus the true parameter value in logspace. The dotted line spanning the diagonal is where perfectly accurate predictions would fall. In the top left of each plot, we show the node features used, and in the bottom right of each plot we show the parameter inferred by the model. The A_{SN1} and A_{SN2} models are successful (R^2 of 0.987 and R^2 of 0.880, respectively), and the A_{AGN} model is unsuccessful (R^2 of 0.102).

model (like shown in the bottom panels of Figure 3), this result signifies that there is some amount of meaningful information regarding the employed DM model of our simulations contained in the structure of the merger events, albeit not as much as with the addition of time dependence.

4.2. WDM Mass Threshold

Our M_{WDM} model was generally accurate in its predictions at low masses, but at greater masses, the accuracy decreased and the uncertainties increased. As shown in the top two panels and bottom left panel of Figure 4, after a mass threshold of ~ 8 keV, the network begins to undershoot the true parameter values consistently, oftentimes making predictions clustered around ~ 10 keV. This result is qualitatively similar to Rose et al. (2023), who train a CNN to infer WDM mass on full box N-body simulations from images with limited success in similar mass ranges (see their Figures 4-6). They attribute this to the resolutions of their simulations, which restricts the maximum WDM mass that is distinguishable. We are similarly constrained by simulation resolution, and so after a certain mass our GNN

cannot easily distinguish between WDM models. In our case, this mass is ~ 8 keV.

4.3. AGN Model Performance

The results from the A_{AGN} model shown in Figure 5 demonstrate that the model cannot infer the AGN feedback parameter, regardless of the node features it learns from. These results match Lin et al. (2024) (see their Figure 2), who train a conditional normalizing flow on 14 node features from individual DREAMS galaxies at $z = 5$ to infer the same parameters as us. Our R^2 values in Section 3.2 are also quantitatively similar to the results in their Section 3.1.

We speculate that the inability of our model to infer A_{AGN} is due to the MW zoom-in simulation suite we have selected for our dataset. The AGN feedback parameter consists of a high-accretion state and a low-accretion state, which are active for different mass ranges. For stellar masses $\lesssim 10^{10.5}$ M$_{\odot}$, the low-accretion state is not active, and therefore does not influence the MW zoom-in suite, as mostly all the subhalos in this suite fall below this mass (Weinberger et al. 2018). Since our dataset is comprised of data features
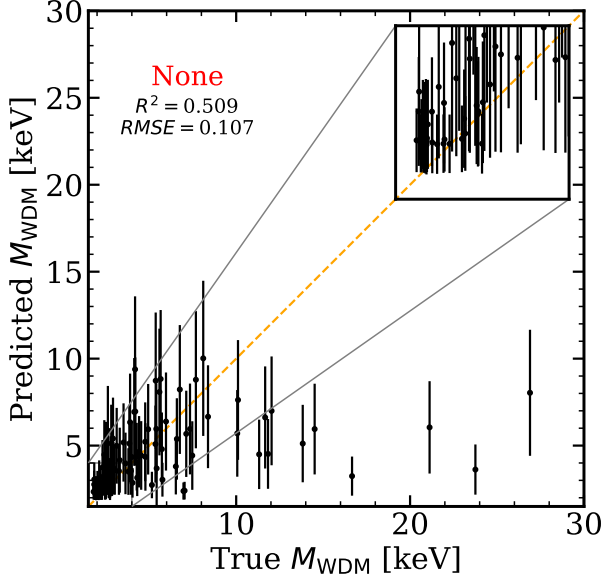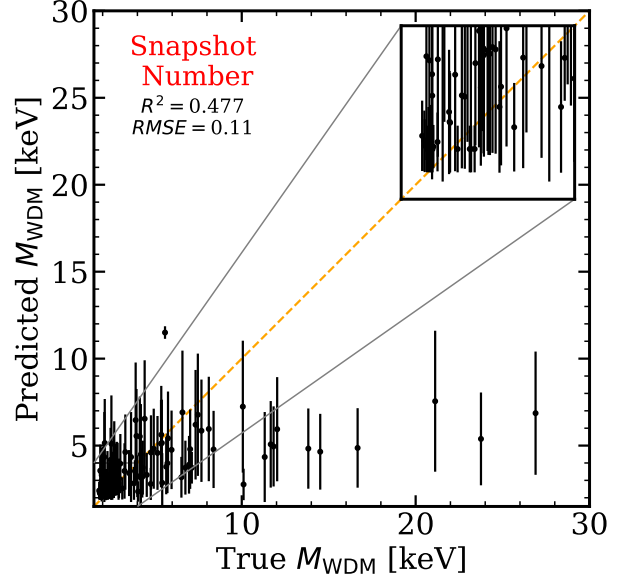
Figure 6. The plotted results of training the WDM GNN without the addition of any node features. The plot shows the inferred WDM particle mass versus the true WDM particle mass. The dotted line across the diagonal represents perfectly accurate predictions. The model makes inferences with a R^2 value of 0.509 and a RMSE of 0.107, which demonstrates a small amount of success.

from these subhalos, there is likely limited information within it regarding A_{AGN} that our GNN can learn from.

4.4. *Pruned Trees*

Results from Jespersen et al. (2022) indicate that reducing the noise of our data could improve the ability of our models to infer parameters. In this section, we experiment with pruning merger trees, that is, removing certain nodes and branches from the trees to make the data less noisy. We try three methods of removing nodes and branches based on three different criteria. The first method is removing all nodes with fewer than 100 DM particles. The second is removing all nodes with less than 100 DM particles that are also not significant in the tree structure (the nodes that are the first and last node in each branch are significant, as these are the nodes that define a branch of the tree). The third is removing all branches that begin with a node with less than 100 DM particles. Between these criteria, the third method produces the simplest and least noisy trees, so we choose to test this method.

We train our WDM GNN with snapshot number as the only node feature from our merger trees pruned using the third method, for 50 trials of 1000 epochs of hyperparameter tuning. We show the results in Figure 7.

Figure 7. The plotted results of training the WDM GNN with pruned merger trees and snapshot number as a node feature. The plot shows the inferred WDM particle mass versus the true WDM particle mass. The dotted line along the diagonal is where perfectly accurate predictions would fall. The model is somewhat successful, with a R^2 value of 0.477 and a RMSE of 0.110, though less so than its unpruned counterpart (shown in Figure 3, top left).

We find that the GNN is able to make accurate predictions, though less accurate than training it on the same node features with our original trees (shown in Figure 4). We see that the model overshoots the true values for many points, until ~ 8 keV, after which all of the points are below the true values. We calculate a R^2 value of 0.454 a RMSE of 0.117, supporting our above findings. The R^2 value is less than the R^2 value from the same node combination without pruning the trees, which indicates that our WDM GNN model benefits from a larger selection of subhalos in our dataset.

4.5. *Sensitivity Tests*

To investigate which specific components of the merger trees are sensitive to WDM, we conduct two tests. First, we train the M_{WDM} GNN with just the most massive progenitors stemming from the first subhalo (called the main branch) of each merger tree, and second with all the nodes in a given tree organized into a single long sequence of edges (flattened trees). We construct flattened trees by trivially drawing edges between consecutive nodes in the data (i.e., the node at index i is connected to the node at index $i+1$) instead of matching through hereditary merger links, which effectively elim-

inates the merger information contained within edges. These tests show if the number of nodes (or, analogously, the number of subhalos) in the tree is more important for WDM mass inference than the edges connecting them. We train the $M_{\rm WDM}$ model for 50 trials of 1000 epochs of hyperparameter tuning with snapshot number as the only node feature for both of these tests, and show the results in Figure 8.

In the left panel, we show the results from training the model on just the main branch. There is a clear non-trend in the predicted values, which indicates that the model is unsuccessful. These results have a of $R^2 = 0.073$ and a RMSE of 0.147, supporting this interpretation. In the right-hand panel, we show the results from training the model on all the nodes, but in a flattened tree structure instead of the normal merger tree. Compared with the results from the original tree with the same node feature (top left panel of Figure 3), the flattened trees qualitatively seem about the same. There still is a trend following the diagonal, particularly at lower masses, and the predicted values drop off after a certain mass threshold. This test has a R^2 value of 0.873 and RMSE of 0.054, and the original tree produced a R^2 of 0.708 and a RMSE of 0.082. The difference in the quantitative metrics between the flattened and original trees does not demonstrate any extreme difference between the two results, as highlighted in the plots. As in Section 3.2, we attribute this difference to variance in the GNN from testing to testing. We repeat these tests for the astrophysics GNNs as well (results shown in Appendix A), and find that the $A_{\rm SN1}$ GNN cannot learn from just the main branch, but can with the flattened tree, similar to the WDM GNN. As the $A_{\rm SN2}$ and $A_{\rm AGN}$ models already cannot learn from the full tree with the snapshot number (shown in Section 3.1), they expectedly are unable to make accurate predictions with either the main branch or flattened tree tests.

These results provide some insight as to where the WDM sensitivities lie within the merger trees. Our tests indicate that the number of subhalos in a merger tree is very sensitive to WDM, whereas the graph structure we use is less so. In comparing Figure 3 (top left panel) and Figure 8 to Figure 4, we also see that the DM halo mass has significant predictive power for the WDM mass. Indeed, previous work has found similar results: the halo mass function and progenitor mass functions have been shown to be sensitive to the free-streaming scale of DM and therefore WDM particle mass (Benson et al. 2013). Though we do not directly measure these mass functions, by including $\mathbf{M_{DM}}$ as a node feature, the merger tree histories naturally gain information regarding these mass functions. This can explain why

$\mathbf{M_{DM}}$ has such predictive power. Across all of our tests, the two strongest sensitivities to WDM mass seem to be the number of subhalos and the DM halo mass.

4.6. *Feature Importance*

We conduct further tests into the WDM sensitivity of each feature using the boosted decision tree based ML algorithm called eXtreme-Gradient Boosting (XG-Boost) (Chen & Guestrin 2016). XGBoost generates an ensemble of decision trees, comprised of a base tree and subsequent trees formed from the residual errors after each iteration. The contributions from these subsequent trees are scaled by the learning rate, and are summed to produce the final prediction from the model. The model is trained to minimize the mean squared error between WDM particle mass predictions and the corresponding true WDM values. Using subsequent learners in this way is more accurate and robust than using a single decision tree for the same purpose.

Our primary interest with this test is evaluating feature importance. Feature importance is a metric of how much a given feature improves the performance of the model. This, combined with the previous tests in Section 4.5, let us compare how two independent ML models highlight WDM sensitivities in our data, giving deeper insight into what the most important components for WDM inference are.

We create our XGBoost model using seven features. Five of these features are the features we used with our GNN ($\mathbf{S}$, $\mathbf{M_{DM}}$, $\mathbf{M_*}$, $\mathbf{M_G}$, $\mathbf{SFR}$). Feature six is, per the results of Section 4.5, the number of nodes $\mathbf{N_{nodes}}$. Feature seven is the ratio of gas mass to DM halo mass $\mathbf{M_G/M_{DM}}$, which we find to have particular significance relative to other ratios and individual features. We compute summary statistics for each feature across a given merger tree, such that the model is given seven data points for each merger tree. We choose to use the 84th percentile as our summary statistic for each feature, which we find to produce more accurate predictions than other summary statistics. We use the same merger tree dataset used for our GNN (described in Section 2.1), though only including subhalos with $z > 7$ and $10^{7.5} {\rm M_\odot} < \mathbf{M_{DM}} < 10^{8.5} {\rm M_\odot}$. We make this restriction to focus on WDM models with the most pronounced free-streaming effects and with the least interference from later-time astrophysical feedback processes. Our dataset is divided with an 80/20 training/testing split.

Our model is trained with six hyperparameters that are tuned using RandomizedSearchCV from the sklearn library (Pedregosa et al. 2012). The hyperparameters are their search space ranges are: the number
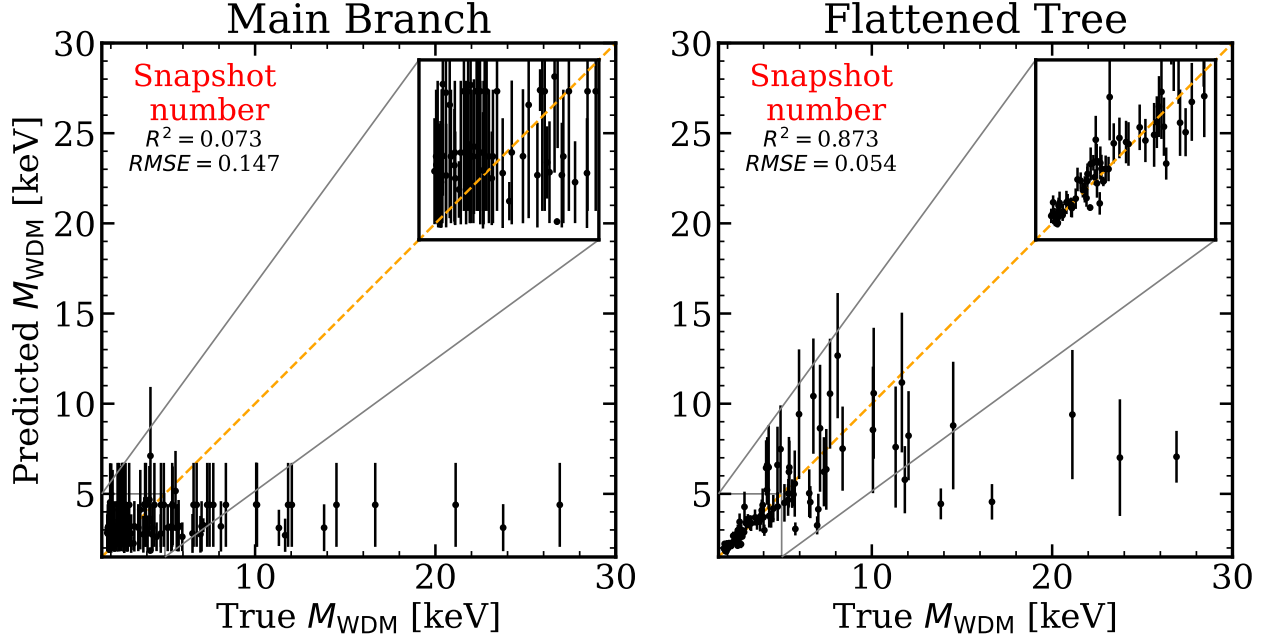
Figure 8. The plotted results of training the WDM GNN with just the main branch (left) and with all the subhalos but only using a single sequence of edges (right). In the first instance, the model fails to infer the WDM particle mass, but in the second, the model does learn to infer accurately, with a R^2 value of 0.873 and a RMSE of 0.054.

of trees (ne) 100 to 600, the depth of each tree (md) 3 to 10, the learning rate (lr) 0.01 to 0.2, the row sampling per tree (ss) 0.4 to 0.6, the feature sampling (fs) 0.4 to 0.6, and the minimum instance weights (mw) 1 to 9. We randomly sample over these ranges for 50 iterations, which finds the best hyperparameters: $ne = 289$, $md = 6$, $lr = 0.019$, $ss = 0.636$, $fs = 0.760$, $mw = 4.297$.

With these hyperparameters, our model makes predictions of the WDM particle mass with a R^2 of 0.847 (successful, though less so than our GNN). We show the feature importance for the model in Figure 9. Clearly, the number of nodes is the most important feature, with an importance of more than double the next feature (0.428). This supports our findings with the GNN tests (Section 4.5)– the number of subhalos in a merger tree is extremely sensitive to WDM. However, unlike with our GNN, the DM halo mass is not important to the XG-Boost model, as it has the lowest feature importance (0.058). It is possible this discrepancy between ML models arises from what specifically they are learning from. Our GNN is trained on each individual subhalo DM mass, whereas the XGBoost model uses summary statistics of the total DM halo mass for a given box. This may indicate that the evolution of DM mass over time, rather than one summary statistic, is much more sensitive to WDM.

This being said, the ratio of gas mass to DM mass seems to be relatively important to the model (fractional importance of 0.123), which is interesting because neither the gas mass nor DM halo mass are important individually. [Ilem] worth talking about? or no

5. CONCLUSIONS

In this paper, we demonstrate that GNNs trained on galaxy merger trees can recover both astrophysical feedback parameters and the underlying DM model properties used within cosmological simulations. We trained several GNNs using an array of combinations of subhalo properties to assess which features were most informative for recovering each model parameter. Our conclusions are as follows:

- We presented a modified version of the COS-MOGRAPHNET architecture designed to take annotated galaxy merger trees as inputs and predict simulation parameters including specifically the adopted WDM particle mass and astrophysics feedback model parameters (Section 2).

- We trained a series of GNNs using data from the DREAMS WDM MW zoom-in suite (Section 3.2) to infer WDM particle masses. Notably, we found that GNNs trained with only time information (snapshot number) were able to predict WDM
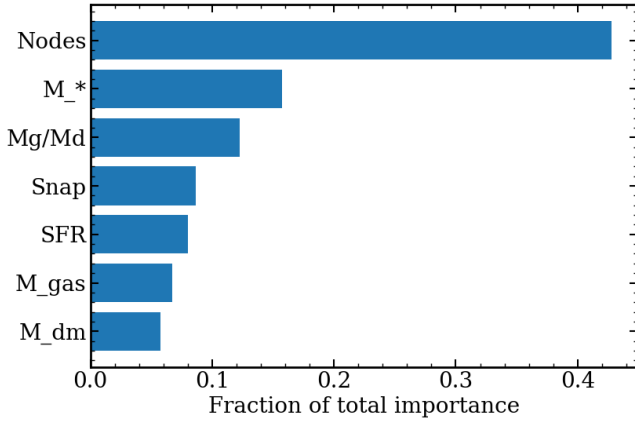
Figure 9. The fraction of total importance for the seven features used to train our XGBoost model. The number of subhalos is more than twice as important as the next most important feature, indicating its extreme sensitivity to WDM. Stellar mass and gas to DM ratio are the next two most important features. The remaining features are marginally important, contributing less than 10 percent each to the total.

masses (Section 3.1 and Figure 3). While this could be attributed to WDM simply impacting the number of nodes within a tree, GNNs trained without time information failed to perform equivalently well (Section 4.1 and Figure 4.1). This indicates that not only the abundance but the temporal structure of the merger tree contains information about the adopted initial matter power spectrum.

- We quantified how progressively enriching the node features (e.g., adding stellar mass, gas mass, DM mass, and star formation rate) improved the network's ability to recover WDM particle masses, reducing prediction scatter and extending the range of WDM masses over which accurate inference could be performed (Figure 4).

- We identified a performance threshold for the WDM inference with GNNs reliably being able to recover the WDM masses in the range $\lesssim$ 6–10 keV but systematically underestimating higher-mass values. We attributed this limitation to the de-

creasingly well-resolved distinction between high-mass WDM models behavior at the resolution of our zoom-in simulations (Sections 3 and 4.2).

- We demonstrated that the same GNN inference can be used to predict the SN wind energy A_{SN1} and the SN wind velocity A_{SN2} astrophysical parameters, but cannot be used to predict the AGN feedback parameter A_{AGN} parameter (Figure 5). We speculate that the performance of our A_{AGN} model is due to the subhalo mass limitations in our MW zoom-in suite, which could potentially be resolved by using a dataset from a different simulation suite (Section 4.3).

- We investigated which components of merger trees are sensitive to WDM. We determined that subhalo abundance in merger trees and the DM halo mass are dominant for WDM inference with our GNN (Sections 4.4, 4.5, Figures 7, 8), whereas the actual graph structure is less so. We directly investigated feature importance using an XGBoost model, again showing subhalo abundance to be highly sensitive (Section 4.6, Figure 9).

This work offers a way to clarify how simulation assumptions shape galaxy properties and their formation histories. This framework can help identify which galaxy features are most sensitive to underlying cosmological and feedback parameters, guiding the design and interpretation of future simulation studies.

DATA ACCESS

The code used to produce the results in this paper is publicly available on GitHub[3].

---

[3] https://github.com/ilemleisher/merger_tree_gnn

REFERENCES

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. 2019, in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19 (New York, NY, USA: Association for Computing Machinery), 2623–2631, doi: 10.1145/3292500.3330701

Battaglia, P., Hamrick, J. B. C., Bapst, V., et al. 2018, arXiv. https://arxiv.org/pdf/1806.01261.pdf

Beck, R., Sadowski, P., Glaser, Y., & Szapudi, I. 2019, in NeurIPS Machine Learning and the Physical Sciences Workshop

Benson, A. J., Farahi, A., Cole, S., et al. 2013, Monthly Notices of the Royal Astronomical Society, 428, 1774

Bode, P., Ostriker, J. P., & Turok, N. 2001, The Astrophysical Journal, 556, 93, doi: 10.1086/321541

Boylan-Kolchin, M., Bullock, J. S., & Kaplinghat, M. 2011, Monthly Notices of the Royal Astronomical Society: Letters, 415, L40, doi: 10.1111/j.1745-3933.2011.01074.x

Chen, T., & Guestrin, C. 2016, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16 (ACM), 785–794, doi: 10.1145/2939672.2939785

Chuang, C.-Y., Jespersen, C. K., Lin, Y.-T., Ho, S., & Genel, S. 2024, The Astrophysical Journal, 965, 101

Dalal, N., & Kochanek, C. S. 2002, The Astrophysical Journal, 572, 25, doi: 10.1086/340303

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, Astrophys. J., 292, 371, doi: 10.1086/163168

de Blok, W. J. G., Walter, F., Brinks, E., et al. 2008, The Astronomical Journal, 136, 2648, doi: 10.1088/0004-6256/136/6/2648

de Santi, N. S., Shao, H., Villaescusa-Navarro, F., et al. 2023a, The Astrophysical Journal, 952, 69

de Santi, N. S., Villaescusa-Navarro, F., Abramo, L. R., et al. 2023b, arXiv preprint arXiv:2310.15234

Dubinski, J., & Carlberg, R. G. 1991, ApJ, 378, 496, doi: 10.1086/170451

Enzi, W., Murgia, R., Newton, O., et al. 2021, Monthly Notices of the Royal Astronomical Society, 506, 5848, doi: 10.1093/mnras/stab1960

Fey, M., & Lenssen, J. E. 2019, Fast Graph Representation Learning with PyTorch Geometric, doi: 10.48550/arXiv.1903.02428

Fitts, A., Boylan-Kolchin, M., Bozek, B., et al. 2019, Monthly Notices of the Royal Astronomical Society, 490, 962

Fluri, J., Kacprzak, T., Refregier, A., Lucchi, A., & Hofmann, T. 2021, Phys. Rev. D, 104, 123526, doi: 10.1103/PhysRevD.104.123526

Garuda, N., Wu, J. F., Nelson, D., & Pillepich, A. 2024, arXiv e-prints, arXiv:2411.12629, doi: 10.48550/arXiv.2411.12629

Gilman, D., Birrer, S., Nierenberg, A., et al. 2019, Monthly Notices of the Royal Astronomical Society, 491, 6077, doi: 10.1093/mnras/stz3480

Hahn, O., & Abel, T. 2011, Monthly Notices of the Royal Astronomical Society, 415, 2101, doi: 10.1111/j.1365-2966.2011.18820.x

Huang, N., Stiskalek, R., Lee, J.-Y., et al. 2025, arXiv preprint arXiv:2507.03707

Iršič, V., Viel, M., Haehnelt, M. G., et al. 2017, Physical Review D, 96, 023522

Jeffrey, N., & Wandelt, B. D. 2020, Solving high-dimensional parameter inference: marginal posterior densities & Moment Networks, doi: 10.48550/arXiv.2011.05991

Jespersen, C. K., Cranmer, M., Melchior, P., et al. 2022, The Astrophysical Journal, 941, 7, doi: 10.3847/1538-4357/ac9b18

Kennedy, R., Frenk, C., Cole, S., & Benson, A. 2014, Monthly Notices of the Royal Astronomical Society, 442, 2487

Klypin, A., Kravtsov, A. V., Valenzuela, O., & Prada, F. 1999, The Astrophysical Journal, 522, 82, doi: 10.1086/307643

Li, H., Xu, Z., Taylor, G., Studer, C., & Goldstein, T. 2018, in Advances in Neural Information Processing Systems, ed. S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett, Vol. 31 (Curran Associates, Inc.). https://proceedings.neurips.cc/paper_files/paper/2018/file/a41b3bb3e6b050b6c9067c67f663b915-Paper.pdf

Lin, S., Villaescusa-Navarro, F., Rose, J., et al. 2024, The Astrophysical Journal, 970, 170, doi: 10.3847/1538-4357/ad54b6

Lovell, M. R., Frenk, C. S., Eke, V. R., et al. 2014, Monthly Notices of the Royal Astronomical Society, 439, 300

Lovell, M. R., Eke, V., Frenk, C. S., et al. 2012, Monthly Notices of the Royal Astronomical Society, 420, 2318

Massara, E., Villaescusa-Navarro, F., & Percival, W. J. 2023, Journal of Cosmology and Astroparticle Physics, 2023, 012

Moore, B. 1994, Nature, 370, 629, doi: 10.1038/370629a0

Moore, B., Ghigna, S., Governato, F., et al. 1999, The Astrophysical Journal, 524, L19, doi: 10.1086/312287

Murgia, R., Merle, A., Viel, M., Totzauer, M., & Schneider, A. 2017, Journal of Cosmology and Astroparticle Physics, 2017, 046

Nadler, E., Drlica-Wagner, A., Bechtol, K., et al. 2021a, Physical review letters, 126, 091101

Nadler, E. O., Birrer, S., Gilman, D., et al. 2021b, The Astrophysical Journal, 917, 7, doi: 10.3847/1538-4357/abf9a3

18

Nelson, D., Pillepich, A., Springel, V., et al. 2019, Monthly Notices of the Royal Astronomical Society, 490, 3234, doi: 10.1093/mnras/stz2306

Oman, K. A., Navarro, J. F., Fattahi, A., et al. 2015, Monthly Notices of the Royal Astronomical Society, 452, 3650, doi: 10.1093/mnras/stv1504

Papastergis, E., Giovanelli, R., Haynes, M. P., & Shankar, F. 2015, A&A, 574, A113, doi: 10.1051/0004-6361/201424909

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2012, CoRR, abs/1201.0490

Pillepich, A., Springel, V., Nelson, D., et al. 2017, Monthly Notices of the Royal Astronomical Society, 473, 4077, doi: 10.1093/mnras/stx2656

Pillepich, A., Nelson, D., Springel, V., et al. 2019, Monthly Notices of the Royal Astronomical Society, 490, 3196, doi: 10.1093/mnras/stz2338

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13, doi: 10.1051/0004-6361/201525830

Ritondale, E., Vegetti, S., Despali, G., et al. 2019, Monthly Notices of the Royal Astronomical Society, 485, 2179

Rodriguez-Gomez, V., Genel, S., Vogelsberger, M., et al. 2015, Monthly Notices of the Royal Astronomical Society, 449, 49, doi: 10.1093/mnras/stv264

Rose, J. C., Torrey, P., Villaescusa-Navarro, F., et al. 2023, Monthly Notices of the Royal Astronomical Society, 527, 739, doi: 10.1093/mnras/stad3260

—. 2025, The Astrophysical Journal, 982, 68, doi: 10.3847/1538-4357/adb8e5

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. 2009, IEEE Transactions on Neural Networks, 20, 61, doi: 10.1109/TNN.2008.2005605

Shao, H., Villaescusa-Navarro, F., Villanueva-Domingo, P., et al. 2023, The Astrophysical Journal, 944, 27

Sobol', I. 1967, USSR Computational Mathematics and Mathematical Physics, 7, 86, doi: https://doi.org/10.1016/0041-5553(67)90144-9

Spergel, D. N., Bean, R., Doré, O., et al. 2007, ApJS, 170, 377, doi: 10.1086/513700

Springel, V. 2010, Monthly Notices of the Royal Astronomical Society, 401, 791, doi: 10.1111/j.1365-2966.2009.15715.x

Springel, V., Pakmor, R., & Weinberger, R. 2019, AREPO: Cosmological magnetohydrodynamical moving-mesh simulation code, Astrophysics Source Code Library, record ascl:1909.010

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

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629, doi: 10.1038/nature03597

Torrey, P., Vogelsberger, M., Genel, S., et al. 2014, MNRAS, 438, 1985, doi: 10.1093/mnras/stt2295

Van Waerbeke, L., Mellier, Y., & Hoekstra, H. 2005, A&A, 429, 75, doi: 10.1051/0004-6361:20041513

Viel, M., Becker, G. D., Bolton, J. S., & Haehnelt, M. G. 2013, Physical Review D—Particles, Fields, Gravitation, and Cosmology, 88, 043502

Villaescusa-Navarro, F., Wandelt, B. D., Anglés-Alcázar, D., et al. 2022a, The Astrophysical Journal, 928, 44, doi: 10.3847/1538-4357/ac54a5

Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., et al. 2021a, The Astrophysical Journal, 915, 71, doi: 10.3847/1538-4357/abf7ba

—. 2021b, Multifield Cosmology with Artificial Intelligence, doi: 10.48550/arXiv.2109.09747

Villaescusa-Navarro, F., Genel, S., Angles-Alcazar, D., et al. 2021c, Robust marginalization of baryonic effects for cosmological inference at the field level, doi: 10.48550/arXiv.2109.10360

Villaescusa-Navarro, F., Genel, S., Anglés-Alcázar, D., et al. 2022b, The Astrophysical Journal Supplement Series, 259, 61, doi: 10.3847/1538-4365/ac5ab0

—. 2023, The Astrophysical Journal Supplement Series, 265, 54, doi: 10.3847/1538-4365/acbf47

Villanueva-Domingo, P. 2022, PabloVD/CosmoGraphNet:v1.0, v1.0, Zenodo, doi: 10.5281/zenodo.6485804

Villanueva-Domingo, P., & Villaescusa-Navarro, F. 2022, The Astrophysical Journal, 937, 115, doi: 10.3847/1538-4357/ac8930

Vogelsberger, M., Genel, S., Sijacki, D., et al. 2013, MNRAS, 436, 3031, doi: 10.1093/mnras/stt1789

Wang, J., & White, S. D. M. 2007, Monthly Notices of the Royal Astronomical Society, 380, 93, doi: 10.1111/j.1365-2966.2007.12053.x

Weinberger, R., Springel, V., & Pakmor, R. 2020, The Astrophysical Journal Supplement Series, 248, 32, doi: 10.3847/1538-4365/ab908c

Weinberger, R., Springel, V., Pakmor, R., et al. 2018, Monthly Notices of the Royal Astronomical Society, 479, 4056, doi: 10.1093/mnras/sty1733

APPENDIX

A. ASTROPHYSICAL PARAMETERS RESULTS EXTENDED

In this section, we show the results for all of the node feature combinations used with the astrophysical parameter models A_{SN1}, A_{SN2}, A_{AGN} in Figure 10.

	S, M$_{\mathbf{G}}$	**S, M$_{\mathbf{G}}$, M$_*$**	**S, M$_{\mathbf{G}}$, M$_*$, M$_{\mathbf{DM}}$**	**S, M$_{\mathbf{G}}$, M$_*$, M$_{\mathbf{DM}}$, SFR**	MB	FT
A_{SN1}	0.737, 0.185	0.966, 0.066	0.953, 0.078	0.987, 0.041	0.021, 0.356	0.703, 0.169
A_{SN2}	−0.003, 0.172	0.541, 0.116	0.837, 0.069	0.880, 0.059	−0.003, 0.172	−0.013, 0.173
A_{AGN}	−0.012, 0.345	−0.013, 0.345	−0.005, 0.344	0.102, 0.325	−0.024, 0.347	−0.012, 0.345

Figure 10. All the results of training the three astrophysical parameter models A_{SN1}, A_{SN2}, A_{AGN} on every combination of node features used, as well as the additional main branch and flattened tree tests (columns 6 and 7, respectively). Each entry in the table shows the calculated coefficient of determination and the root mean squared error, respectively.