Graphical representations of galaxies and dark matter halos

John F. Wu
Space Telescope Science Institute
Johns Hopkins University

https://jwuphysics.github.io

1. Introduction

Galaxies are collections of stars, gas, and dark matter formed via hierarchical merging. The complex dynamics of internal galaxy processes and environmental effects give rise to the diverse array of galaxy properties observed throughout the Universe. In order to better understand how galaxies grow and evolve, we must study both their host dark matter halos as well as their large-scale environments.

Galaxies have small sizes (~kpc) relative to their typical separations (~Mpc), and they interact with their surroundings in multiple ways, e.g., gravitational attraction, tides, and ram pressure.

Mathematical graphs are well-suited for representing galaxies in the cosmic web, as graphs enforce useful inductive biases, e.g., equivariance to permutations and invariance under the E(3) group action [1].

Recent works have shown that graph neural networks (GNNs) trained on simulated galaxy samples can be used to estimate galaxy merger histories [2][3], large-scale alignments [4], dark matter halo mass of centrals [5], and cosmological parameters [6].

In this work, we show that *all central and satellite subhalo masses* can be jointly learned via cosmic GNNs using galaxy positions, velocities, and stellar masses.

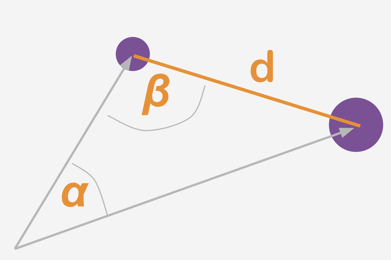
Our simulated galaxy sample is based on the $z = 0$ SUBFIND catalogs from the Illustris TNG300-1 cosmological box [7]. We ensure that $M_{halo} > 10^{11}\,M_{sun}$, $M_\star > 10^{9.5}\,M_{sun}$, and $N_{\star\,particles} > 50$.

In order to construct the graphs (*see below*), we split the data into sub-boxes with ~50 Mpc per side. Note that the sub-boxes are padded by half the linking length to keep graphs independent.

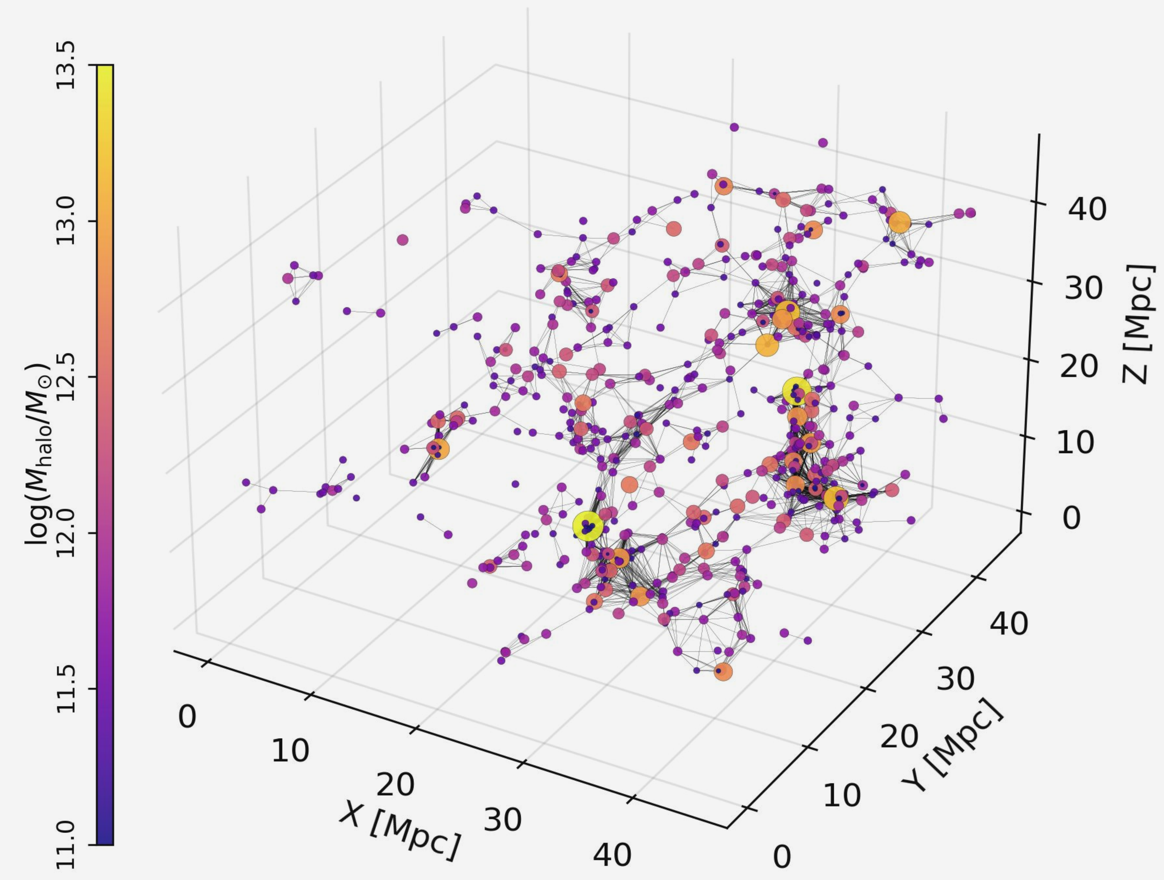
2. Inferring dark matter halo masses with cosmic graphs

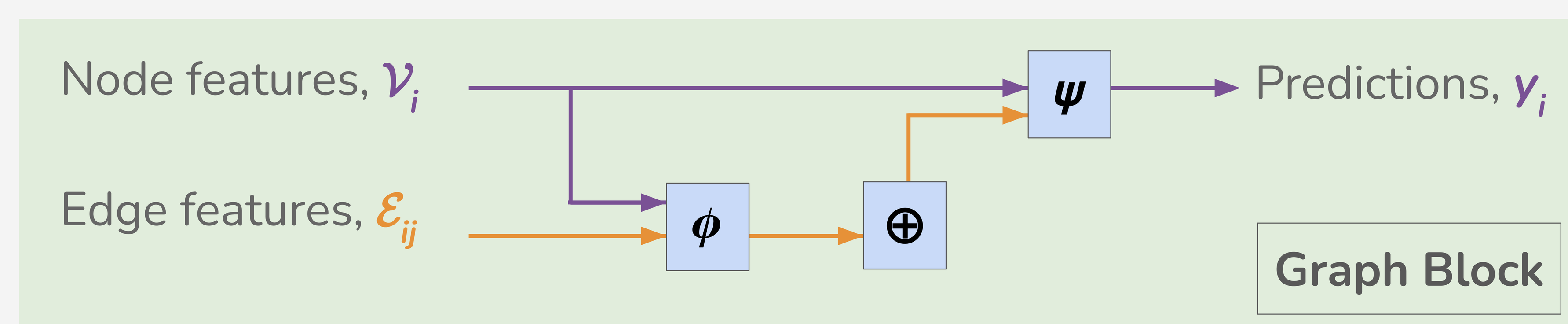
Node features, $\mathcal{V}$:
$\{x, y, z, v_x, v_y, v_z, M_\star\}$

Edge features, $\mathcal{E}$:
$\{d, \alpha, \beta\}$



Predictions, y_i:
$\{M_{halo}\}$





ϕ ψ are multi-layer perceptrons that have two hidden layers, each followed by LayerNorm and a ReLU activation.

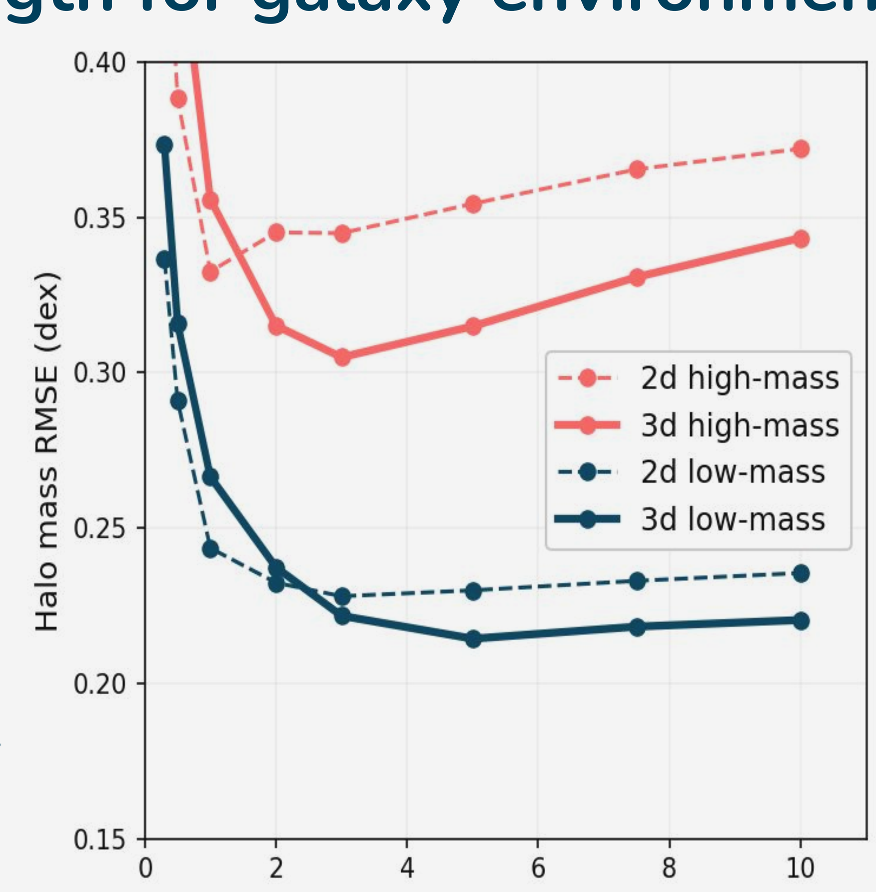
$\oplus$ is an aggregation function that concatenates *max*, *sum*, and *mean*-pooled activations, over all edges ij that connect nodes j to i.

We train GNNs consisting of one **Graph Block** for 1000 epochs, and evaluate results using 6-fold cross-validation. This method is similar to [6].

3. A characteristic linking length for galaxy environments

We turn off self-loops, which forces the GNN to estimate halo mass purely from neighboring galaxy properties, rather than from a combination of **halo-galaxy** and **halo-environment** connections.

Results are shown for galaxies with $M_\star > 10^{10.5}\,M_{sun}$ and $M_\star \leq 10^{10.5}\,M_{sun}$.



The lowest RMSE for halo mass predictions occurs at **3 - 5 Mpc** scales for 3d GNNs (*solid lines*).

In general, GNNs that predict halo masses benefit from information on scales *larger than the dark matter halo virial radius* (> 1 Mpc).

At very large scales, we begin to (*i*) crop out information due to our sub-box padding, and (*ii*) add noise because of irrelevant information.

We find that the lowest RMSE occurs at **1 - 3 Mpc** scales for 2*d* GNNs (*dashed lines*), i.e., when we only provide projected positions **x**, **y**, and line-of-sight velocity $\mathbf{v_z}$.

For the smallest scales, we note that the GNN in 2*d* projection *outperforms* the 3*d* GNN! This is because the extra environmental information from projected large scale structure outweighs the contaminating structures!

4. Towards GNN predictions for observed galaxies

Our results show that projected galaxy structures, some of which may be contaminants, can actually *improve* halo mass predictions in a small "pencil beam" field. For reference, a single JWST NIRCam field at $z = 0.5$ spans only 800 kpc.

We use full simulation catalogs for each sub-box rather than group catalogs, which is more realistic and applicable to observations; group membership catalogs leak information about the virial mass.

We found a characteristic linking length of several Mpc, which is consistent with observations of synchronized galaxy properties on large scales (e.g., galactic conformity; [8][9]). However, additional work is required to better understand the nature of this characteristic scale.

References and acknowledgments

[1] Battaglia et al., 2018, *arXiv*, 1806.01261
[2] Jespersen et al., 2022, *ApJ*, 941, 7
[3] Jagvaral et al., 2022, *MNRAS*, 516, 2406
[4] Tang & Ting, 2022, *arXiv*, 2207.02786
[5] Villanueva-Domingo et al., 2022, *ApJ*, 935, 30
[6] Villanueva-Domingo & Villaescusa-Navarro, 2022, *ApJ*, 937, 115
[7] Pillepich et al., 2018, *MNRAS*, 473, 4077
[8] Olsen et al., 2021, *ApJ*, 913, 45
[9] Ayromlou et al., 2023, *MNRAS*, 519, 1913

I am grateful for many invaluable discussions with Christian Kragh Jespersen. Valuable conversations with Peter Behroozi, and Tjitske Starkenburg, Paco Villaescusa-Navarro, and Risa Wechsler during the *galevo23* KITP Program also helped steer this work. This research was supported in part by the National Science Foundation under Grant No. NSF PHY-1748958.