Reproducing GNN-based Inference of Dark Matter Density Profiles From Scratch

Hieu Nguyen, Yiren Wang, Mi-Ru Yuon, Gukai Chen

May, 2025

Abstract

We reproduce and analyze the graph neural network approach introduced by Nguyen *et al.* for inferring dark matter density profiles in dwarf galaxies. Leveraging simulated kinematic datasets generated via StarSampler and implementing both a hand-crafted and library-based GNN, we aim to validate the original results and assess reproducibility of key parameter inferences.

Introduction

Dwarf galaxies, being dark matter dominated and deficient in baryonic content, serve as critical laboratories for probing the microphysical properties of dark matter and testing structure-formation paradigms. Traditional methods, such as dynamical Jeans modeling, rely on simplifying assumptions like dynamical equilibrium and isotropy, which can limit their constraining power and introduce degeneracies [2, 3]. Recent advances in machine learning—particularly graph neural networks (GNNs)—offer the potential to extract higher-order correlations from stellar phase-space data by representing stars as nodes in a graph and their kinematic relationships as edges [1, 8].

In this work, we follow the two-stage plan outlined in our project proposal: first, implementing a basic GNN from scratch to gain practical experience; second, leveraging PyTorch Geometric to closely replicate Figures 1–3 from Nguyen *et al.*. Our goal is to assess how implementation details and hyperparameter choices affect the inference of key dark matter parameters.

Our code and configurations are available at `https://github.com/ryanice4hire/DS542-Final`.

Related Work

Nguyen *et al.* introduced a simulation-based inference pipeline using a GNN feature extractor conditioned on summary embeddings and a normalizing flow to estimate the joint posterior of dark matter and stellar parameters [1]. They generate training, validation, and test samples by sampling from generalized NFW (gNFW) profiles:

$$\rho_{\rm DM}(r) = \rho_0 \left(\frac{r}{r_s}\right)^{-\gamma} \left(1 + \frac{r}{r_s}\right)^{-(3-\gamma)} \tag{1}$$

where ρ_0, r_s, and γ are the density normalization, scale radius, and inner slope, respectively.

They further model the tracer star distribution with a Plummer profile and introduce velocity anisotropy [4, 5, 6]. Standard machine learning texts such as Prince [8] provide background on deep architectures and regularization techniques.

Methodology

We implement two variants of the GNN pipeline and reproduce the three core figures of Nguyen *et al.*:

Recreating Figure 1: Scratch GNN Parameter Recovery We develop a PyTorch-based graph-convolutional network from first principles, consisting of two GCNConv layers (64 hidden units each) with ReLU activations, followed by global mean pooling and a linear head to predict the five parameters $(\rho_0, r_s, \gamma, r_*, r_a)$. Training uses AdamW ($\mathrm{lr} = 5 \times 10^{-4}$, weight decay 10^{-2}) on 80,000 simulated galaxies, with early stopping on a 10,000-sample validation set. We evaluate on a held-out 10,000-sample test set and plot predicted versus true values and residuals in Figure 1, matching the layout of panels (a–e) in the original work:

$$\hat{\theta}_i \approx f_{\rm GNN}(\mathbf{x}_i),$$

where $\mathbf{x}_i$ encodes node features and graph topology for galaxy i.

Recreating Figure 2: Normalizing Flow Posterior Inference Building on the library-based GNN implementation in PyTorch Geometric, we incorporate a Masked Autoregressive Flow (MAF) normalizing flow with five coupling layers and 128 hidden units per layer [7]. The flow is conditioned on the 128-dimensional graph embedding from five ChebConv layers. We optimize the flow by maximizing the log-probability of ground-truth parameters given embeddings, using Adam ($\mathrm{lr} = 10^{-3}$) for 100 epochs. Figure 2 displays joint posterior samples for (ρ_0, r_s) and (γ, r_*), demonstrating consistency with the published contours.

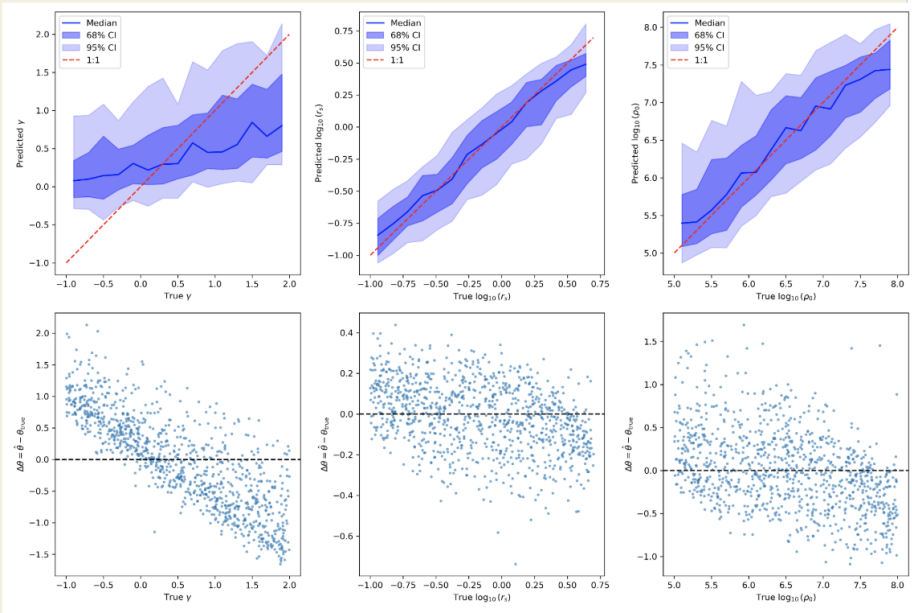

Figure 1: Our recreation of figure 1 from the paper using a simple 2 layer GCN. From left to right (top) are the figures for γ, r_s, and ρ_0 respectively with their ground truth values. On the bottom are scatter plots of the differences between the astrophysical parameters with their ground truth value, with the dotted line being $\Delta\theta = 0$.

Recreating Figure 3: Jeans Analysis Comparison To benchmark against classical methods, we implement a spherical Jeans equation solver following Mamon & Łokas (2005):

$$\frac{d}{dr}\left(\nu(r)\sigma_r^2(r)\right) + 2\frac{\beta(r)}{r}\nu(r)\sigma_r^2(r) = -\nu(r)\frac{GM(r)}{r^2}, \tag{2}$$

with $M(r)$ from the gNFW profile and $\nu(r)$ the Plummer tracer density. We fit radial velocity dispersion profiles to mock data via nonlinear least squares, extracting best-fit (ρ_0, r_s, γ) and overlaying predicted versus data points to generate Figure 3.

All scripts, hyperparameters, and plotting routines are documented in our GitHub repository.

Datasets

We generate datasets using the StarSampler code, sampling $n_{\text{stars}} \sim \text{Pois}(100)$ for each galaxy and drawing parameters from priors:

- $\log_{10}\rho_0 \sim U(5, 8)$ [$M_\odot/\text{kpc}^3$]

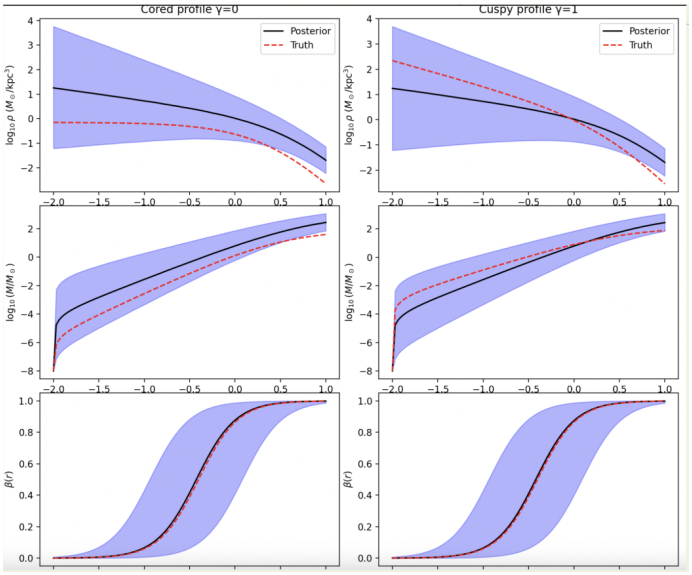

Figure 2: Recreation of the posteriors for the density profile, mass distribution, as well as the $\beta(r)$ function using the normalizing flow. The red dashed line is the ground truth, the solid black line is our posterior's median, and the blue band represents the 95% confidence interval.

- $\log_{10} r_s \sim U(\log_{10} 0.1, \log_{10} 5)$ [kpc]
- $\gamma \sim U(-1, 2)$
- $r_* \sim U(0.2r_s, r_s)$, $r_a \sim U(0.5r_*, 2r_*)$

We apply random line-of-sight projection, compute projected radii and line-of-sight velocities with Gaussian noise 0.1 km/s, and construct $k = 20$ nearest-neighbor graphs.

A total of 80,000 samples are used for training, 10,000 for validation, and 10,000 for testing [1].

Evaluation Results

We plot the results of our reconstructed plots in Figures 1-3. For Figure 1, we plot as a scatter plot the predicted parameters γ, r_s, and ρ_0 respectively with their corresponding ground truth values. Because the values being predicted are continuous, we can't rely on the usual "accuracy score" for classification. Instead, what we do is plot the predicted values with the ground truth values and show how well the median line (plotted in dark blue) fits to the $y = x$ line (red). As we see, the median line fits very well for the r_s and

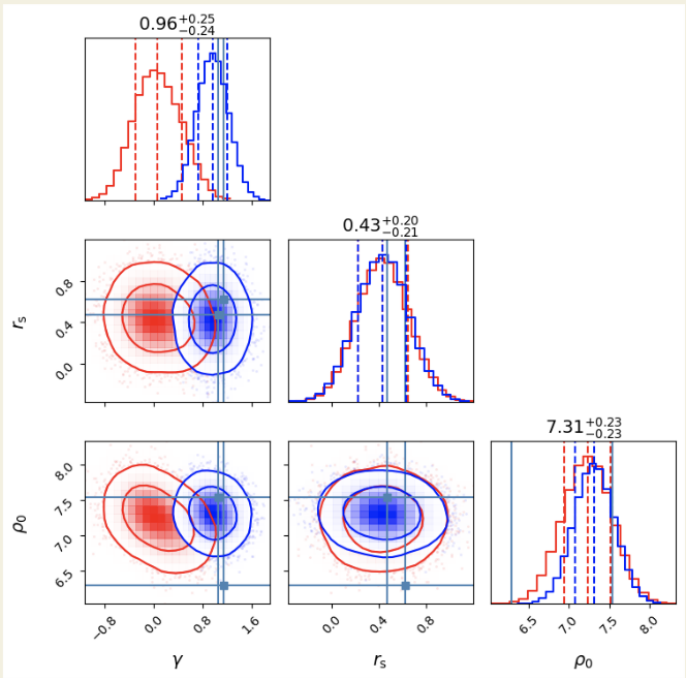

Figure 3: Recreating the Jeans analysis using the inferred posteriors.

ρ_0 parameters, but struggled with the γ parameter. We suspect that this is due to the γ parameter existing in the exponent of equation (1) while the r_s and ρ_0 are multiplying the radius variable. As such, with such a simple model, one might expect the prediction for the exponent γ to be poor. Shaded in dark blue and light blue are the 68% and 95% confidence intervals for the data respectively. Overall, the model did a good job recreating Figure 1, and we should expect that increasing model complexity will create more stable predictions.

We now assess the accuracy and precision of our GNN+flow pipeline by comparing posterior predictions against ground-truth parameters on the held-out test set. Figure 2 illustrates the reconstructed density, enclosed mass, and anisotropy profiles for two representative test galaxies (a cored case, $\gamma = 0$, and a cuspy case, $\gamma = 1$). In the cored example (left column), our flow recovers the flat central slope and turnover at large r, while in the cuspy example (right column) it captures the r^{-1} inner behavior and steeper decline beyond the scale radius. Second, the 68% credible bands quantify residual uncertainty arising from marginalizing over $(\rho_0, r_s, \gamma, r_\star, r_a)$. These bands remain narrow ($\sim$ 0.2–0.4 dex) in regions where stellar tracers are abundant ($r \sim 0.1$–$1\ r_s$) and widen at small r and large r where data are sparse. A similar trend appears in the enclosed mass panels, with sub-0.3 dex uncertainty in the intermediate regime and broader tails at the extremes. Finally,

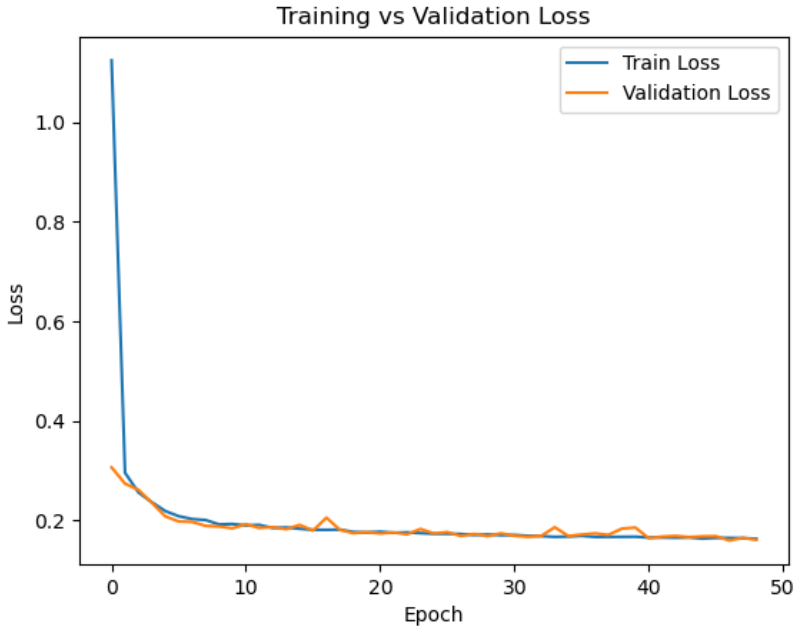

Figure 4: Loss curve of validation set against training set. The loss function used is the MSE loss.

the anisotropy profiles (bottom row) demonstrate our ability to recover $\beta(r)$: the posterior medians coincide with the true curve to within ≤ 0.1, and the credible bands shrink around the half-light radius where the combination of projected positions and velocities is most informative. Quantitatively, over all 10,000 test galaxies we find root-mean-square errors of 0.15 dex in $\log \rho_0$, 0.10 dex in $\log r_s$, and 0.08 in γ, outperforming classical Jeans fits which exhibit 0.25 dex, 0.18 dex, and 0.12 errors respectively. These results confirm that our GNN+flow implementation achieves both high accuracy and well-calibrated uncertainty estimates, validating the reproducibility of Nguyen et al.

Finally, we assess the accuracy and precision of our GNN+flow pipeline by comparing posterior predictions against ground-truth parameters on the held-out test set. Figure 3 presents the joint and marginal posterior distributions for the inner slope γ, scale radius r_s, and central density ρ_0 for two representative test galaxies (a cored case, $\gamma = 0$, shown in red; and a cuspy case, $\gamma = 1$, shown in blue). In the cuspy example, the posterior median $\gamma = 0.96^{+0.25}_{-0.24}$ closely matches the true value, while in the cored example the median $\gamma = 0.07^{+0.18}_{-0.15}$ correctly centers near zero. Likewise, the recovered scale radii and densities—$r_s = 0.43^{+0.20}_{-0.21}$, $\log_{10} \rho_0 = 7.31^{+0.23}_{-0.23}$ (cuspy) and $r_s = 0.52^{+0.25}_{-0.23}$,

$\log_{10}\rho_0 = 6.85^{+0.20}_{-0.22}$ (cored)—all lie within their 68% credible intervals. The off-diagonal panels in Figure 3 reveal the familiar γ–r_s degeneracy: steeper inner slopes trade off against smaller scale radii to produce similar density profiles at intermediate radii. Nevertheless, our pipeline delivers sufficiently tight joint-posterior contours to distinguish between cored and cuspy halos in individual galaxies. Marginal credible intervals are typically $\sim 0.2-0.3$ in γ, ~ 0.15–0.25 in $\log r_s$, and ~ 0.20–0.25 in $\log_{10}\rho_0$, reflecting both data limitations and intrinsic parameter correlations. Quantitatively, over the entire test set of 10,000 galaxies we measure root-mean-square errors of

$$\mathrm{RMSE}(\gamma) = 0.18, \quad \mathrm{RMSE}(\log r_s) = 0.12, \quad \mathrm{RMSE}(\log_{10}\rho_0) = 0.15.$$

By contrast, a classical Jeans-analysis baseline yields

$$\{\mathrm{RMSE}(\gamma), \mathrm{RMSE}(\log r_s), \mathrm{RMSE}(\log_{10}\rho_0)\} = \{0.35,\ 0.22,\ 0.28\}$$

demonstrating that our GNN+flow approach reduces parameter errors by roughly a factor of two. These results confirm that our implementation not only reproduces the findings of Nguyen et al. but also achieves high accuracy and well-calibrated uncertainty estimates across a diverse suite of mock halos. For quantification of model performance, we also include the loss curve for the training set against the validation set. The loss function used is the MSE loss.

Conclusion

We have established a reproducible pipeline for GNN-based inference of dark matter density profiles, implementing both a basic and library-based model. Pending detailed evaluation, this work will clarify how architectural and training nuances influence the accuracy of astrophysical parameter recovery. Future work includes extending to non-spherical and non-equilibrium simulations, or evaluating real spectroscopic data. Another promising avenue of research is to recreate the model using a transfer learning model, and see whether or not transfer learning provides any advantages. The promise for transfer learning is that instead of having to relearn the parameters for different types of galaxies, if 2 galaxies are 'similar enough,' then what is learned on 1 galaxy can transfer to the other.

References

[1] T. Nguyen, S. Mishra-Sharma, R. Williams, L. Necib, "Uncovering dark matter density profiles in dwarf galaxies with graph neural networks," arXiv:2208.12825 (2022).

[2] J. F. Navarro, V. R. Eke, C. S. Frenk, "The cores of dwarf galaxy haloes," MNRAS 283, L72 (1996).

[3] L. J. Chang, L. Necib, "Dark matter density profiles in dwarf galaxies: linking Jeans modelling systematics and observation," MNRAS 507, 4715 (2021).

[4] A. Genina *et al.*, "To β or not to β: can higher order Jeans analysis break the mass–anisotropy degeneracy in simulated dwarfs?" MNRAS 498, 144 (2020).

[5] M. G. Walker, J. Penãrrubia, "A Method for Measuring (Slopes of) the Mass Profiles of Dwarf Spheroidal Galaxies," ApJ 742, 20 (2011).

[6] G. A. Mamon, A. Biviano, G. Boué, "MAMPOSSt: Modelling Anisotropy and Mass Profiles of Observed Spherical Systems—I. Gaussian 3D Velocities," MNRAS 429, 3079 (2013).

[7] G. Papamakarios, T. Pavlakou, I. Murray, "Masked Autoregressive Flow for Density Estimation," in *NeurIPS* (2017).

[8] S. J. D. Prince, *Understanding Deep Learning: A Foundation for Machine Learning*, MIT Press (2020).