Uncovering dark matter density profiles in dwarf galaxies with graph neural networks

Tri Nguyen,[1,2,*] Siddharth Mishra-Sharma,[1,3,4] Reuel Williams,[5] and Lina Necib[1,2]

[1]*The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Cambridge, Massachusetts 02139, USA*

[2]*Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA*

[3]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

[4]*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*

[5]*Department of Mathematics, Princeton University, Princeton, New Jersey 08544, USA*

(Received 16 September 2022; accepted 6 January 2023; published 10 February 2023)

Dwarf galaxies are small, dark-matter-dominated galaxies, some of which are embedded within the Milky Way. Their lack of baryonic matter (e.g., stars and gas) makes them perfect test beds for probing the properties of dark matter—understanding the spatial dark matter distribution in these systems can be used to constrain microphysical dark matter interactions that influence the formation and evolution of structures in our Universe. We introduce a new method that leverages simulation-based inference and graph-based machine learning in order to infer the dark matter density profiles of dwarf galaxies from observable kinematics of stars gravitationally bound to these systems. Our approach aims to address some of the limitations of established methods based on dynamical Jeans modeling. We show that this novel method can place stronger constraints on dark matter profiles and, consequently, has the potential to weigh in on some of the ongoing puzzles associated with the small-scale structure of dark matter halos, such as the core-cusp discrepancy.

DOI: 10.1103/PhysRevD.107.043015

I. INTRODUCTION

Cosmological structure formation is known to proceed hierarchically—smaller structures seed the formation of larger structures [1]. Dark matter (DM) plays an outsized role in this process, acting as a "scaffolding" on which structure evolution plays out. At the same time, the precise mechanism of structure formation is keenly sensitive to the microphysical properties of DM, e.g., the nature of its self-interactions. Deviations from the canonical Λ cold dark matter paradigm of cosmology would be imprinted in the properties of DM clumps (known as "halos") on smaller spatial scales. Robustly characterizing the distribution of small-scale structures in our Universe may therefore hold the key to answering one of the major unsolved questions in particle physics and cosmology—the particle nature of DM.

Dwarf galaxies are small galaxies, some of which are embedded within larger galaxies like the Milky Way. They are dominated by DM [2], making them versatile astrophysical laboratories for DM studies. A major goal in cosmology and particle physics is to detect nongravitational interactions of DM. For the canonical weakly interacting massive particle (WIMP) DM paradigm, one of the main avenues to do so is DM indirect detection: WIMPs could annihilate into Standard Model (SM) particles, producing striking signatures from DM-overdense regions in γ-ray observations [3–12]. Being deficient in baryonic matter, dwarf galaxies act as ideal targets for indirect detection, with a relatively large predicted ratio of DM signal to astrophysical background.

A pervasive puzzle in cosmology is the so-called core-cusp discrepancy, referring to the question of whether the inner DM density profiles of dwarf galaxies are cuspy (steeply rising) or cored (flattened) [13,14]. N-body simulations using Λ cold dark matter cosmology suggest that in the absence of baryonic physics, cold DM halos follow the cuspy Navarro-Frenk-White density (NFW) profile [15], which is characterized by a steep rise in the density $\rho \propto r^{-1}$ at small halo-centric radii r. However, recent measurements of stellar dynamics suggest that these systems could instead have a flattened density profile at their center, also known as a core [16,17]; see Ref. [18] for a review. Potential solutions to the core-cusp discrepancy range from stellar feedback that ejects baryons and flattens the DM central density profile [19–22] to alternative DM models like self-interactions [23–25].

DM density profiles in dwarf galaxies are traditionally inferred using spectroscopic observations of line-of-sight

*tnguy@mit.edu

velocities and angular positions of stars gravitationally bound to these systems. In particular, integral moments of the Jeans equation can be used to relate the velocity dispersions of tracer stars to the gravitational potential of the system [26,27]. Although Jeans modeling has proven highly successful for modeling DM distributions in dwarf galaxies, there are several caveats and limitations associated with this approach (see, e.g., [28–30]). For example, Jeans modeling assumes the system is in dynamical equilibrium, which may not be a robust assumption given the active merger history of the Milky Way (see Ref. [31] for a review). Assumptions such as the isotropy of the gravitating system are also often necessary in order to enable a tractable analysis. Finally, by relying on a simplified description of the data through second moments of the stellar velocity distribution, inference based on Jeans modeling is likely to lose some of the salient information available in observations. Absent additional assumptions, this is expressed as a degeneracy between the mass profile of the system and the anisotropy structure of stellar orbits, known as the mass-anisotropy degeneracy. Exploiting additional information about the modeled stellar phase-space distribution can further inform the DM density profile, helping break this degeneracy. Several methods have been proposed in the literature to this end [32], including using higher-order moments of the Jeans equation [29,33,34], leveraging multiple distinct tracer populations [35–37], including measured proper motions when these are available [38,39], and alternative strategies to solving the Jeans equation [40].

This paper introduces a new machine learning-based approach for linking observed stellar properties to the DM density profiles of dwarf galaxies. Our method is based on forward modeling simulated dwarf galaxy systems and corresponding observations, learning to summarize representative features from these datasets using graph neural networks, and performing simulation-based inference to simultaneously extract the spatial profiles associated with the DM and stellar components of the dwarf galaxy.

II. METHODOLOGY

We describe, in turn, the forward model used in this study and its realization via simulations, the representation of stellar kinematic data as a graph, and finally the feature-extractor graph neural network and simulation-based inference procedure.

A. Datasets and the forward model

In this proof-of-principle exposition, we consider idealized simulations of spherical and dynamically equilibrated dwarf galaxies. Our forward model is fully specified by the joint distribution function (DF) $f(\vec{x},\vec{v})$ of positions and velocities of stars following a certain (*a priori* unknown) spatial distribution (known as the "light profile"). These tracer stars are gravitationally bound to a DM halo with a density profile we wish to infer. We use the public code StarSampler[1] to generate simulated realizations of stellar kinematics (6D position and velocity phase-space components) from the forward model. StarSampler uses importance sampling [41–43] to sample the DF of positions and velocities of tracer stars in a given DM potential.

We model the DM profile using the generalized NFW (GNFW) profile [15],

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

which depends on three free parameters: the density normalization ρ_0, the scale radius r_s, and the asymptotic inner slope γ. $\gamma = 1$ corresponds to a cuspy NFW profile, while $\gamma = 0$ corresponds to a pure DM core. We consider these two parameter points as benchmarks, since the ability to robustly distinguish between the two possibilities would offer a path towards resolution of the core-cusp discrepancy.

We assume a stellar density distribution $\nu(r)$ that follows a 3D Plummer profile [44]:

$$\nu(r) = \frac{3L}{4\pi r_\star^3}\left(1+\frac{r^2}{r_\star^2}\right)^{-5/2}, \tag{2}$$

where L is the total luminosity and $r_\star$ is the scale length. We also introduce a velocity anisotropy profile $\beta(r)$ in order to model deviations from circular orbits; $\beta(r)$ is defined similarly to Refs. [45,46] as

$$\beta(r) = \frac{r^2}{r^2+r_a^2}, \tag{3}$$

which has an additional parameter r_a describing the radius of transition from isotropic velocity orbits at small radii to radially biased orbits at larger radii.

In total, our model has three DM parameters (ρ_0, r_s, γ) and two stellar parameters $(r_\star, r_a)$. Assuming the gravitational potential of the system is dominated by DM, and the model is independent of L in Eq. (2). We provide further details of the forward model, phase-space distribution function, and a summary of the prior distributions of the DM and stellar parameters in Appendix A.

We generate 80 000 training samples, 10 000 validation samples, and 10 000 test samples using the prior parameter distributions. Each sample contains the 3D positions and 3D velocities of tracer stars with respect to the center of a dwarf galaxy. The number of stars in each galaxy is sampled from a Poisson distribution $n_{\rm stars} \sim {\rm Pois}(\mu_{\rm stars})$. We set $\mu_{\rm stars} = 100$ stars in our baseline benchmark, roughly corresponding in order of magnitude to the number of stars typically observed in dwarf galaxies of interest [47–50]. We explore variations on this choice in Appendix B 1.

[1] https://github.com/maoshenl/StarSampler.

B. Data preprocessing and graph construction

For each kinematic sample, we randomly draw a line-of-sight axis and project the galaxy onto the 2D plane perpendicular to it. We then derive the 2D projected spatial coordinates with respect to the center of the galaxy (X, Y) and line-of-sight velocities $v_{\rm los}$ for each star in these coordinates, reflecting typically available observations. To study the validity of the method before the inclusion of large measurement errors, we assume a Gaussian velocity noise model with standard deviation 0.1 km/s. The effect of larger measurement errors is shown in Appendix B 1.

We can represent the stellar kinematic data in the form of a potentially weighted, undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$, where $\mathcal{V}$ is a set of nodes representing $|\mathcal{V}| = N_{\rm stars}$ individual stars, $\mathcal{E}$ is a set of edges, and $A \in \mathbb{R}^{N_{\rm stars} \times N_{\rm stars}}$ is an adjacency matrix describing the weights of connections between vertices. This representation is well suited for our purposes since the stars in a dwarf galaxy have no intrinsic ordering, and the graph structure can efficiently capture information about the phase-space correlation structure, including higher-order moments [51].

In our analysis, each node represents a star, with the node features being its line-of-sight velocity $\tilde{v}_{\rm los}$ and the projected radius $R = \sqrt{X^2 + Y^2}$. We choose to use R instead of the full (X, Y) coordinates in order to incorporate projective rotational invariance into the graph representation, which was found to enhance the simulation efficiency of our method.

To determine the graph edges $\mathcal{E}$, we calculate pair-wise distances between all stars using (X, Y), then connect each star to its k-nearest stars including itself. Since the edges are assumed to be undirected, each star can be connected to more than k other stars. We found $k = 20$ to provide a good trade-off between model performance and computational overhead. Finally, we do not include edge weights in our graph, however, we have experimented with a variety of weighting schemes, including attention-based learned weights [52], and found them to perform similarly in downstream inference to the unweighted case.

C. Neural network architecture and optimization

We use a graph neural network (GNN) $g_\varphi \colon \mathcal{G} \to \mathbb{R}^{N_{\rm feat}}$ in order to extract $N_{\rm feat}$ summary features from the constructed graph representation $x \in \mathcal{G}$ of mock dwarf galaxy stellar kinematic data. Here φ represents the parameters of the graph neural network. The feature-extraction network consists of five graph-convolutional layers, each with 128 channels, based on convolutions in the Fourier domain using a basis of Chebyshev polynomials of order 4 as filters [53]. This is followed by a global mean pooling layer which aggregates the permutation-equivariant features into a permutation-invariant representation, and a fully connected layer that projects the output onto a set of $N_{\rm feat} = 128$ summaries $g_\varphi(x)$. We explore variations on the graph-convolution scheme in Appendix B.

The joint posterior $\hat{p}_\phi(\theta|g_\varphi(x))$ of the five parameters of interest θ characterizing the DM and stellar profiles is modeled using a normalizing flow [54–56]—a class of flexible generative models that allow for efficient density estimation as well as sampling. The flow transformation (with parameters ϕ) is conditioned on the summary features extracted by the graph neural network and its negative log-density $-\log \hat{p}_\phi(\theta|g_\varphi(x))$ is used as the optimization loss. Our flow model consists of four masked autoregressive flow transformations, each using a two-layer masked autoencoder for distribution estimation with hidden dimension 128 [56,57]. This method falls under the class of approaches known as simulation-based inference (see Ref. [58] for a review), specifically neural conditional posterior estimation [59,60].

The GNN and normalizing flow parameters $\{\varphi, \phi\}$ are optimized simultaneously on the 80 000 simulated training samples using the AdamW optimizer [61,62] with an initial learning rate of 5×10^{-4} and a weight decay of 10^{-2} using a batch size of 64. At the end of each epoch, we evaluate the loss on the 10 000 held out validation samples and reduce the learning rate by a factor of 10 if no improvement is seen after four epochs. We stop training if the validation loss has not improved after 10 epochs. Model training typically terminates after ∼30–40 epochs, which takes ∼30 minutes on a single NVIDIA Tesla V100 GPU.

III. RESULTS AND CONCLUSIONS

We apply our pipeline to 10 000 test dwarf galaxies and summarize our results in Fig. 1. For each galaxy, we condition the trained normalizing flow on features extracted using the trained GNN feature extractor and draw 10 000 samples from the joint DM and stellar posterior. We then compute the marginal medians as the predicted parameters and sort them into bins based on their truth values. Figure 1 shows the median (solid blue line), middle-68% (blue bands), and middle-95% (light blue bands) credible intervals for each bin of the DM parameters. In general, our method successfully recovers individual DM parameters consistent with the underlying truth.

To demonstrate the ability of our method to distinguish between a cored ($\gamma = 0$) and cuspy ($\gamma = 1$) DM profile, in Figs. 2 and 3 we show the inferred posteriors on two test dwarf galaxies with the same DM density normalization ($\rho_0 = 10^7\ {\rm M}_\odot/{\rm kpc}^3$), scale radius ($r_s = 1$ kpc), and stellar profile, but with different inner density slopes ($\gamma = 0$ and $\gamma = 1$).

Figure 2 shows the posteriors on the recovered density profile (top row), enclosed mass profile (middle row), and orbital anisotropy profile (bottom row) as a function of halocentric radius of a cored (left) and a cuspy (right) profile. The middle-68% and middle-95% credible intervals are

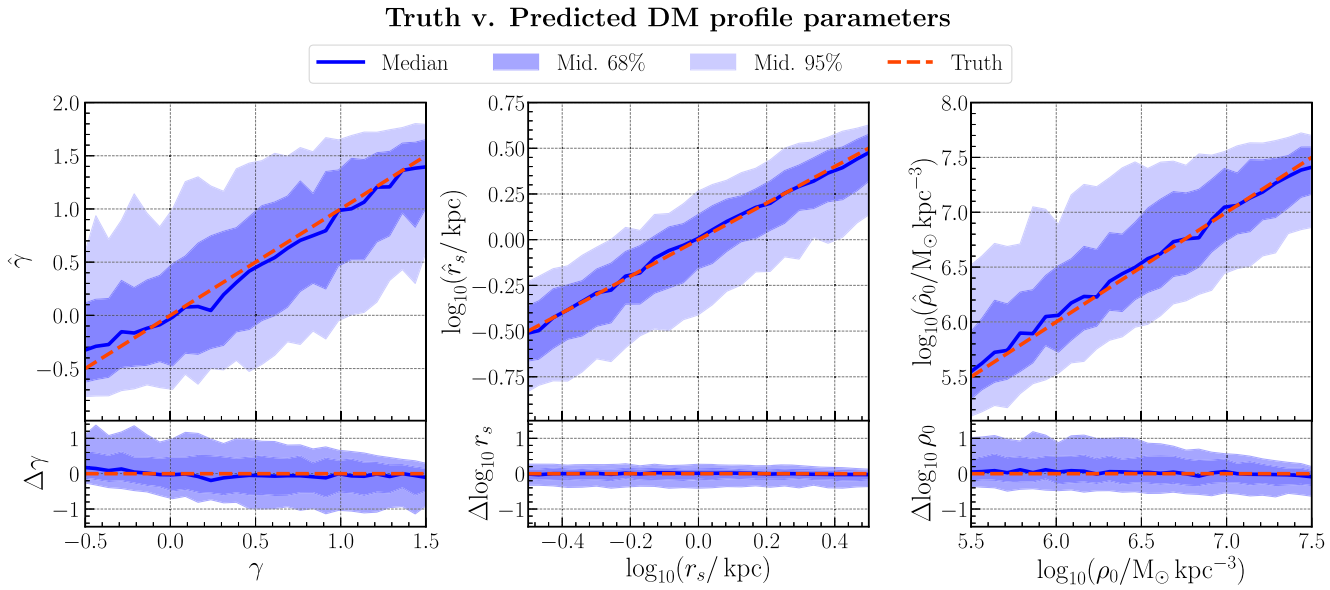

FIG. 1. A comparison between the predicted and the truth values of the DM parameters on 10 000 test galaxies for our baseline scenario. For each galaxy, the predicted parameters are taken to be the marginal medians of the joint posterior binned in truth values. The median (solid blue line), middle-68% percentile (dark blue band), and middle-95% (light blue band) containment regions of each bin are shown. The diagonal dashed red line denotes where the predicted and truth values are equal. The bottom row shows the prediction error on the median $\Delta\theta \equiv \hat{\theta} - \theta_{\mathrm{truth}}$.

shown as blue bands, and the truth profiles are shown with the dashed red lines. We find our method is able to successfully reconstruct the density, mass, and orbital anisotropy profiles at small and large radii.

We apply Jeans analysis on these two test galaxies using a procedure similar to that used in Ref. [30]. The posterior is approximated with nested sampling [63,64] using the module DYNESTY [65] with $n_{\mathrm{live}} = 500$ live points and a convergence tolerance of $\Delta \ln Z = 0.1$ on the estimated evidence. Details of the Jeans analysis procedure are provided in Appendix A 2. Figure 3 shows the joint and marginal DM posteriors from Jeans modeling (left panel) and our method (right panel) for $\gamma = 0$ (red) and $\gamma = 1$ (blue), with the middle-68% and middle-95% credible intervals as the contour lines. In the Jeans analysis, we see significant overlap between the γ posteriors of the cored and cuspy halos. On the other hand, γ posteriors inferred by our method have substantially smaller overlap. See Appendix B 2 for a quantitative comparison of the separability.

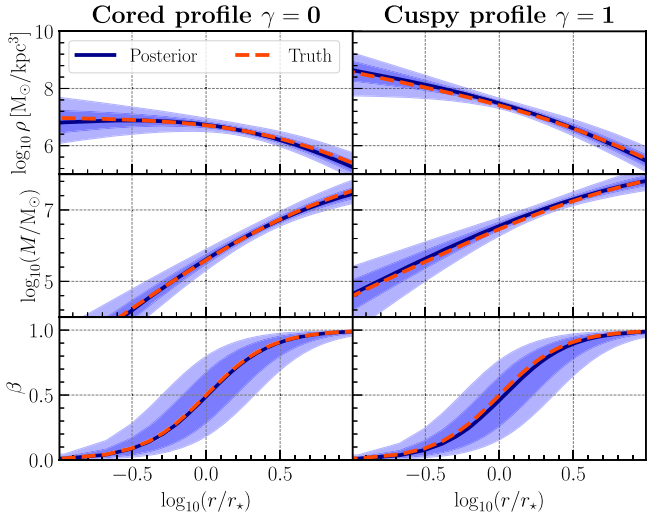

FIG. 2. Example inferred posteriors of the density profile (top row), enclosed mass profile (middle row), and velocity anisotropy profile (bottom row) for dwarf galaxies with a cored DM profile (left) and a cuspy DM profile (right). The dashed red line is the truth profile, while the blue line and bands represent the median, middle-68%, and middle-95% credible intervals, respectively.

An important quantity in indirect searches for DM is the astrophysical J factor—the integral along the line-of-sight s and over solid angle Ω of the squared DM density corresponding to a source target,

$$J = \int \mathrm{d}s \int \mathrm{d}\Omega\, \rho^2(s, \Omega), \tag{4}$$

accurately determining it is therefore important for robustly interpreting results of DM indirect detection experiments using dwarf galaxies as targets [3–12].

In Fig. 4, we show the inferred J factors, normalized to the truth values, for 100 dwarf galaxies randomly sampled from our test dataset using our method (red error bars) and compare these to the corresponding J factors obtained using Jeans analysis (green error bars). The middle panel of Fig. 4 shows ratios of uncertainties on the $\log_{10} J$ factors

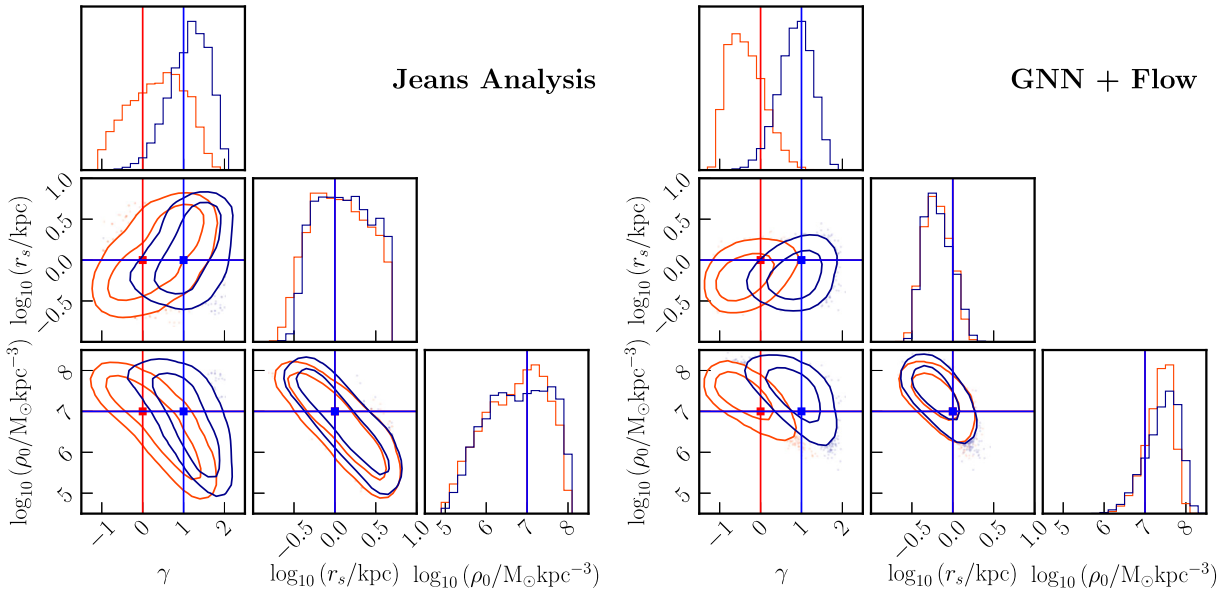

FIG. 3. Example corner plots of the posterior DM parameters from Jeans dynamical modeling (left) and our method (right) on two test galaxies with cored DM profile (red) and cuspy DM profile (blue). Both galaxies have the same central slope ρ_0 and scale radius r_s. The contour lines show the 68% and the 95% credible intervals. It can be seen that our method provides a stronger constraint on the DM parameters and is able to distinguish more cleanly between a cored and cuspy profile.

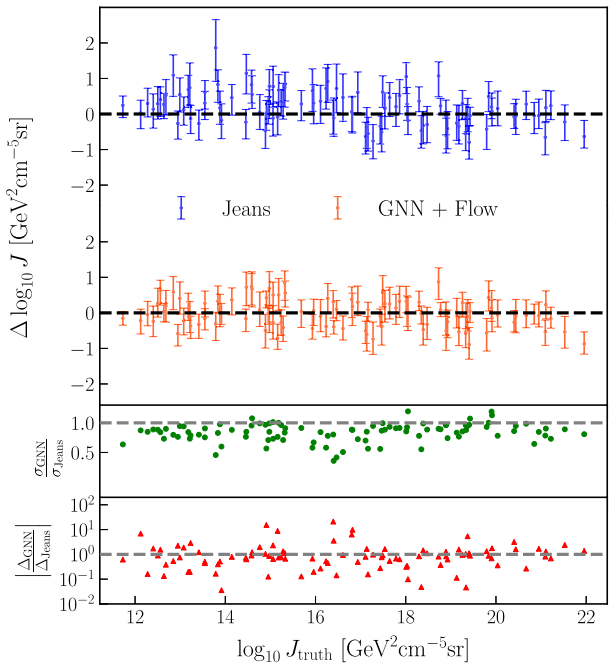

FIG. 4. Comparison between the J factors predicted by the Jeans analysis and our method. Top: differences between the predicted and true $\log_{10} J$ factors for the Jeans analysis (blue) and our method (red). Middle: ratios of the symmetrized errors on the $\log_{10} J$ factors, $\sigma_{\log_{10} J}$. Bottom: ratios of the absolute differences between the truth and predicted $\log_{10} J$ factors. Our method can accurately recover the J-factor posteriors while generically providing tighter constraints than the traditional Jeans analysis.

obtained by our method to those obtained by the Jeans analysis. It can be seen that our method generically provides more stringent constraints than the Jeans analysis, as expected from the fact that the DM density parameters are overall better constrained. For the cases studied, this log-uncertainty is on average ∼20% and up to a factor of ∼2 smaller for our method than that obtained using Jeans analysis. The bottom panel shows the ratios of the absolute differences between the truth and predicted $\log_{10} J$ factors obtained by our method as compared to those obtained by the Jeans analysis. The scatter around unity indicates that neither method is systematically biased, as expected, and serves as an additional cross-check of our analysis. Details of the J-factor calculation and additional results comparing the two methods are shown in Appendices A 3 and B 3.

To conclude, this paper introduces a novel method to reconstruct the DM density profiles of Milky Way dwarf galaxies from measured kinematics of tracer stars based on graph neural networks and simulation-based inference. The method compares favorably with and can outperform established approaches based on Jeans dynamical modeling in speed, flexibility, as well as constraining power. The latter is due to the fact that our method incorporates more information about the phase-space structure of bound stars, contrasted with Jeans-based methods which typically rely on the second moments of the velocity distribution. Additionally, the method simultaneously models the stellar light profile and does not require fitting it beforehand. Although in this paper we used simulations of orbitally anisotropic spherical systems in order to enable a direct comparison with existing methods, a particular strength of

our method is the ability to incorporate nonequilibrium dynamics using cosmologically realistic simulations of isolated dwarfs as well as satellites of Milky Way–like systems, accounting for baryonic effects like tidal disruption [66] and supernova feedback [25]. We defer this extension, as well as application of our method to observational data, to future work.

Code used to reproduce the results of this paper is available at [67].

ACKNOWLEDGMENTS

We thank Laura Chang, Mariangela Lisanti, and Linda Xu for helpful conversations. This work was performed in part at the Aspen Center for Physics, which is supported by National Science Foundation Grant No. PHY-1607611. This work is supported by the National Science Foundation under Cooperative Agreement No. PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions [68]). This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics of U.S. Department of Energy under Grant Contract No. DE-SC0012567. R. W. acknowledges support from the Office of Undergraduate Research at Princeton University and is grateful to Mariangela Lisanti for guidance. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation Grant No. ACI-1548562. Specifically, it used the Bridges-2 system, which is supported by NSF Award No. ACI-1928147, at the Pittsburgh Supercomputing Center (PSC) [69,70]. This research made use of the BILBY [71], DYNESTY [65], IPython [72], Jupyter [73], MATPLOTLIB [74], NFLOWS [75], NumPy [76], PyTorch [77], PyTorch Geometric [78], PyTorch Lightning [79], and SciPy [80] software packages.

APPENDIX A: ADDITIONAL DETAILS ON THE ANALYSIS

In Appendix A we describe additional details of our analyses, including the assumed stellar phase-space distribution function in Appendix A 1, the Jeans dynamical modeling method in Appendix A 2, computation of the annihilation J factor in Appendix A 3, and the assumed prior distributions in Appendix A 4. In Appendix B we show additional results of our analysis, including systematic variations on the baseline analysis in Appendix B 1, a quantitative comparison of the inner density slope and J-factor posteriors between our method and the Jeans analysis in Appendices B 2 and B 3, respectively, tests of statistical coverage in Appendix B 4, and sensitivity of our method to observational projection in Appendix B 5.

We elaborate here on several details of the analysis presented in the paper. We show a schematic illustration of our method, including a rough breakdown of the different steps of the pipeline, in Fig. 5.

1. Details on the forward model and phase-space distribution function

In this section, we describe in detail the forward model used to generate the dataset. We model the DF using the Osipkov-Merritt model as proposed by Refs. [45,46]. The Osipkov-Merritt model depends on the angular momentum J and relative energy per unit mass $\mathcal{E} = \phi - v^2/2$, where ϕ is the gravitational potential and v is the velocity, through the variable $Q \equiv \mathcal{E} - J^2/(2r_a)$. Here, r_a is the scale radius of the velocity anisotropy profile $\beta(r) = r^2/(r^2 + r_a^2)$ that defines the transition from an isotropic velocity dispersion at small radii to a radially biased dispersion at larger radii. To solve for the DF $f(\mathcal{E}, J) = f(Q)$, we first integrate $f(Q)$ over the velocity space to obtain the stellar mass-density profile

$$\rho_\star(r) = \frac{4\pi}{1 + r^2/r_a^2} \int_0^\phi \mathrm{d}Q f(Q) \sqrt{2(\phi - Q)}. \tag{A1}$$

Note that $f(Q) = 0$ for $Q < 0$. The final DF $f(Q)$ can be obtained by Abel transforming $\rho_\star$,

$$\begin{aligned} f(Q) &= \frac{1}{2\pi\sqrt{2}} \frac{\mathrm{d}G(Q)}{\mathrm{d}Q}, \\ \text{where } G(Q) &= -\int_0^Q \frac{\mathrm{d}\rho_Q}{\mathrm{d}\phi} \frac{\mathrm{d}\phi}{\sqrt{Q - \phi}} \\ \text{and } \quad \rho(Q) &= \left(1 + \frac{r^2}{r_a^2}\right) \rho_\star(r). \end{aligned} \tag{A2}$$

Assuming that the system is dominated by DM, the gravitational potential ϕ depends on the DM density profile via Poisson's equation $\nabla^2\phi = -4\pi G \rho_{\mathrm{DM}}(r)$. The DM density profile is parametrized through the generalized Navarro-Frenk-White model as in the main text:

$$\rho_{\mathrm{DM}}^{\mathrm{GNFW}}(r) = \rho_0 \left(\frac{r}{r_s}\right)^{-\gamma} \left(1 + \frac{r}{r_s}\right)^{-(3-\gamma)}, \tag{A3}$$

where ρ_0 is the density normalization, r_s is the DM scale radius, and γ is the inner density slope. The stellar density profile is

$$\rho_\star(r) = \rho_{0,\star} \left[1 + \frac{r^2}{r_\star^2}\right]^{-5/2}, \tag{A4}$$

where $r_\star$ is the stellar scale radius and $\rho_{0,\star}$ is the density normalization of the stellar profile. For $\rho_{0,\star} = 3M/4\pi r_\star^3$ and $M \propto L$, where M and L is the enclosed mass and luminosity, this is proportional to the 3D Plummer profile

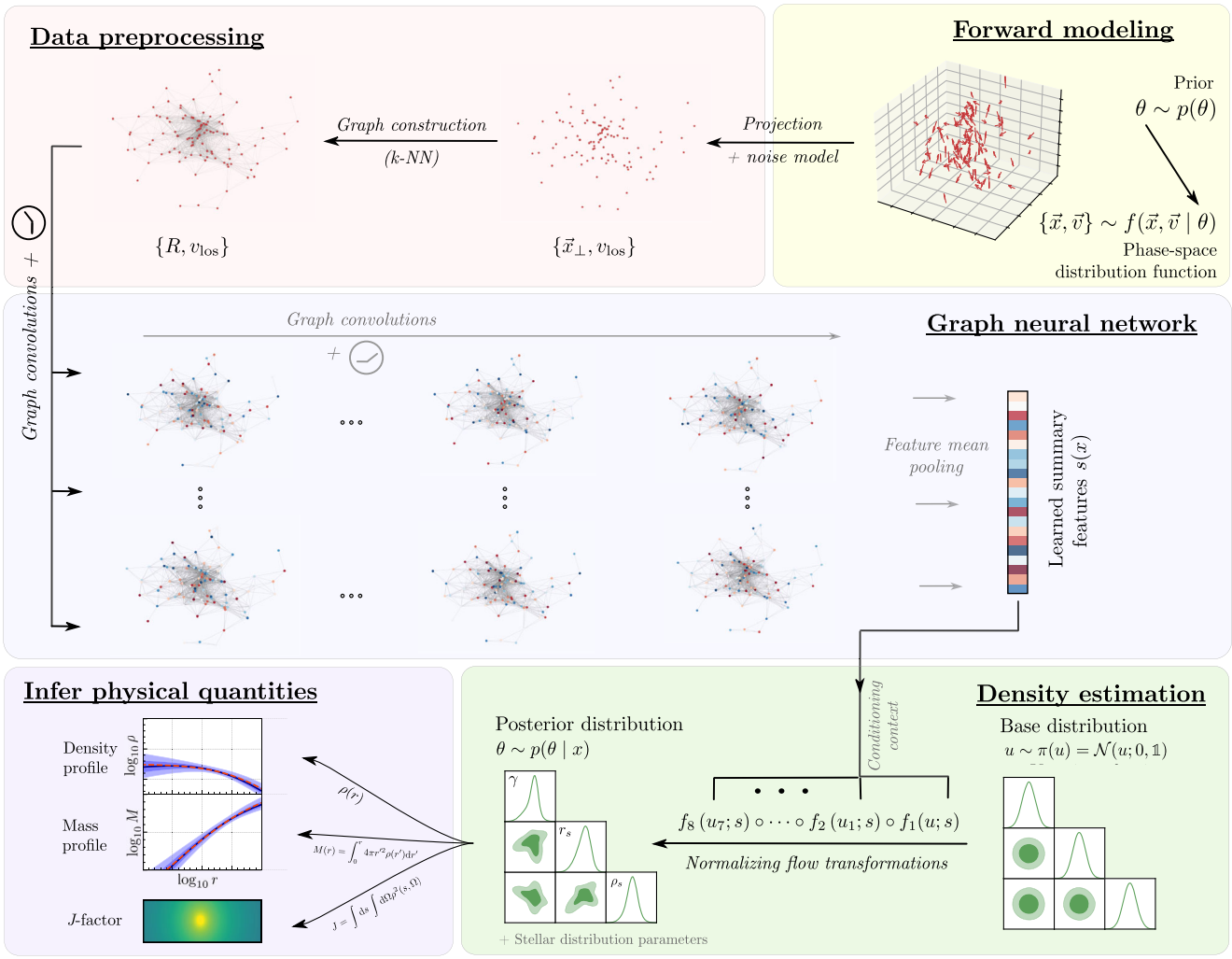

FIG. 5. A schematic illustration of our method for inferring DM density profiles of dwarf galaxies from observed stellar kinematics.

$$\nu(r) = \frac{3L}{4\pi r_\star^3}\left(1 + \frac{r^2}{r_\star^2}\right)^{-5/2}, \tag{A5}$$

which we use to characterize our stellar profile in the main text. The DF $f(Q)$ may be written in terms of the radius r and the tangential and radial velocity v_r and v_t [i.e., $f(Q) = f(r, v_r, v_t)$] since $J = rv_t$ and $\mathcal{E} = \phi(r) - (v_t^2 + v_r^2)/2$. StarSampler samples the coordinates r, v_r, v_t of each star from $f(r, v_r, v_t)$ using importance sampling [41–43]. The 6D coordinates (x, y, z, v_x, v_y, v_z) are then calculated by assuming a random projection direction and spherical symmetry.

2. Details on Jeans analysis procedure

In this section, we briefly outline the Jeans analysis procedure used in this work, closely following Ref. [30]. Following the derivation from Refs. [81,82], we first assume the system follows the collisionless Boltzmann equation

$$\frac{\partial f}{\partial t} + \vec{v}\frac{\partial f}{\partial \vec{x}} - \frac{\partial \Phi}{\partial \vec{x}} \cdot \frac{\partial f}{\partial \vec{v}} = 0, \tag{A6}$$

where ϕ is the gravitational potential, $f = f(\vec{x}, \vec{v})$ is the phase-space distribution function, and $(\vec{x}, \vec{v})$ are the phase space coordinates of tracer stars.

Working in the dwarf galaxy's spherical coordinate system (r, θ, ϕ) and assuming spherical symmetry and steady state, we multiply Eq. (A6) by the radial velocity v_r and integrate over all velocity components to obtain the spherical Jeans equation

$$\frac{1}{\nu}\left[\frac{\partial}{\partial r}(\nu\sigma_r^2) + \frac{2\beta(r)}{r}(\nu\sigma_r^2)\right] = -\frac{\partial \phi}{\partial r} = -\frac{GM(<r)}{r^2}, \tag{A7}$$

where $\nu = \int \mathrm{d}^3\vec{v} f(\vec{x}, \vec{v})$ is the number density of the tracer stars, σ_i is the velocity dispersion $\sigma_i = \sqrt{\langle v_i^2 \rangle - \langle v_i \rangle}$ for $i \in (r, \theta, \phi)$, and $\beta(r) = 1 - (\sigma_\theta^2 + \sigma_\phi^2)/(2\sigma_r^2)$ is the velocity anisotropy profile. The gravitational potential ϕ is

assumed to be dominated by DM and may be written as $\phi = -GM(<r)/r$, where G is the gravitational constant and $M(<r)$ is the enclosed mass of DM. The Jeans equation (A7) has therefore the solution

$$\sigma^2(r)\nu(r) = \frac{1}{g(r)}\int_r^\infty \frac{GM(<r')\nu(r')}{r'^2} g(r')\mathrm{d}r'$$
$$\text{where } g(r) = \exp\left(2\int_0^r \frac{\beta(r')}{r'}\mathrm{d}r'\right). \tag{A8}$$

Projecting Eq. (A8) along the line of sight using the Abel transformation $s(r) \to S(R)$ for the spherically symmetric function $s(r)$,

$$S(R) = 2\int_R^\infty \frac{s(r)r\mathrm{d}r}{\sqrt{r^2-R^2}}, \tag{A9}$$

we obtain

$$\sigma_p^2(R)I(R) = 2\int_R^\infty \left(1-\beta(r)\frac{R^2}{r^2}\right)\frac{\nu(r)\sigma_r^2(r)r}{\sqrt{r^2-R^2}}\mathrm{d}r, \tag{A10}$$

where σ_p is the projected velocity dispersion profile and $I(R)$ is the projected number density of tracer stars, also known as the surface brightness or the light profile. In our analysis, we parametrize the DM density profile using the GNFW profile in Eq. (A3). The light profile $I(R)$ is the projection along the line of sight of the 3D Plummer profile in Eq. (A5) and is given by

$$I(R) = \frac{L}{\pi r_\star^2}\left(1+\frac{R^2}{r_\star^2}\right)^{-2}, \tag{A11}$$

which has two free parameters: the total luminosity L and the scale length $r_\star$.

The Jeans analysis requires two separate fits: fitting the light profile and subsequently fitting the DM profile. We first fit L and $r_\star$ using the priors in Table I. We approximate the posteriors as Gaussian distributions and use them as the priors for L and $r_\star$ in the second fit. To compare the performance between Jeans analysis and our machine learning framework, the DM priors are set to be the same as those used to generate the training datasets for the GNN and normalizing flow method, and are summarized in Table I.

For the initial fit of the light profile, we bin the data in $\log_{10}$-spaced bins in the projected radius R. The number of bins is chosen to be $\sim\sqrt{N_{\rm stars}}$, where $N_{\rm stars}$ is the number of stars. Similarly to Ref. [30], we assume Poisson uncertainties on the number of stars in each bin. Let $n_i(\theta)$ and $\hat{n}_i$ be the predicted and mean number of stars (with θ is the parameters of the light profile model) in ith bins. We construct the binned likelihood [83]

TABLE I. Prior ranges for the DM and stellar parameter for the Jeans analysis and our method.

	Parameter	Prior distribution
DM density profile	$\log_{10}(\rho_0/(\mathrm{M}_\odot\,\mathrm{kpc}^{-3})$	$\mathcal{U}(5,8)$
	$\log_{10}(r_s/\mathrm{kpc}))$	$\mathcal{U}(-1,0.7)$
	γ	$\mathcal{U}(-1,2)$
Light profile (GNN + Flows)	$r_\star/r_s$	$\mathcal{U}(0.2,1)$
	$r_a/r_\star$	$\mathcal{U}(0.5,2)$
Light profile (Jeans analysis)	$\log_{10}(L/\mathrm{L}_\odot)$	$\mathcal{U}(-2,5)$
	$\log_{10}(r_\star/\mathrm{kpc})$	$\mathcal{U}(-3,3)$
	$r_a/r_\star$	$\mathcal{U}(0.5,2)$
	$\bar{v}/(\mathrm{km\,s}^{-1})$	$\mathcal{U}(-100,100)$

$$\ln\mathcal{L}_{\rm Plummer} = -\frac{1}{2}\sum_{i=1}^{N_{\rm stars}}\frac{(\hat{n}_i - n_i(\theta))^2}{V_i - V_i'(\hat{n}_i - n_i(\theta))}, \tag{A12}$$

where $V = \sigma_{\rm lo}\sigma_{\rm hi}$ and $V' = \sigma_{\rm hi} - \sigma_{\rm lo}$; here, $\sigma_{\rm lo}$ and $\sigma_{\rm hi}$ are the asymmetric Poisson uncertainties. We refer to Ref. [30] for further details. As described in the main text, we sample the posterior with nested sampling [63,64] using the DYNESTY module [65]. We use $n_{\rm live} = 500$ live points and a convergence tolerance of $\Delta\ln\mathcal{Z} = 0.1$ on the estimated remaining contribution to the log-evidence. In Fig. 6, we show the initial Plummer fit for the two test galaxies presented in the result of the main text. The posteriors for L and $r_\star$ are well constrained and agree well with the true Plummer profile.

As mentioned, we approximate the posteriors of L and $r_\star$ as Gaussian distributions and use them as priors in the second fit of the DM profile. Unlike the initial Plummer fit, for the Jeans analysis, we construct an unbinned Gaussian likelihood. The likelihood is given by Ref. [84],

$$\mathcal{L}_{\rm Jeans} = \prod_{i=1}^{N_{\rm stars}}\frac{(2\pi)^{-1/2}}{\sqrt{\sigma_p^2(R_i)+\Delta v_i{}^2}}\exp\left[-\frac{1}{2}\left(\frac{(v_i-\bar{v})^2}{\sigma_p^2(R_i)+\Delta v_i{}^2}\right)\right], \tag{A13}$$

where $\sigma_p(R)$ is the projected velocity dispersion profile in Eq. (A10), $\bar{v}$ is the mean velocity of tracer stars, and v_i and Δv_i is the line-of-sight velocity and its measurement error for star i. We treat the mean velocity $\bar{v}$ as a nuisance parameters and fit it together with the DM and light profile, using the prior distribution in Table I. We sample the joint DM and light profile posteriors using DYNESTY [65] with the same sampler configuration as in the initial Plummer fit.

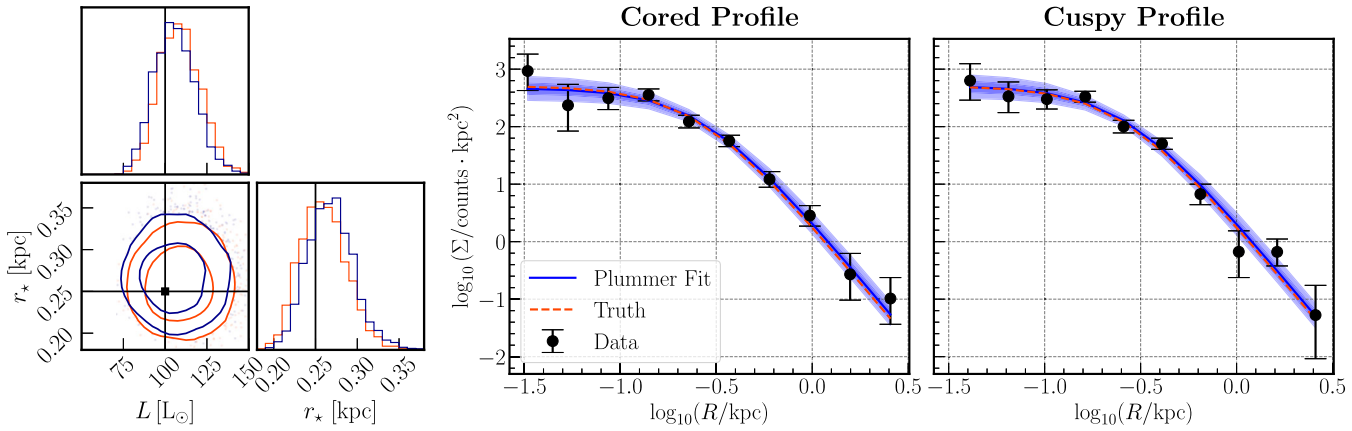

FIG. 6. The initial Plummer fit for the two test galaxies presented in the result of the main text. The left panel shows the 68% and 95% credible intervals of the posterior distributions of L and $r_\star$. Both galaxies have the same light profile ($L = 100$ L$_\odot$ and $r_\star = 0.25$ kpc) but different DM profiles: cored DM profile (red) versus cuspy DM profile (blue). The right panel shows the best fit, the truth Plummer profile, and the binned data with Poisson uncertainties for both galaxies. The dark blue and light blue bands show the middle-68% and 95% credible intervals of the reconstructed profile.

3. Details on the J-factor calculation

In this section, we outline the details on the annihilation J-factor calculation. The J factor is defined as the integral along the line-of-sight s and over solid angle Ω of the squared DM density corresponding to a source target,

$$J = \int \mathrm{d}s \int \mathrm{d}\Omega \rho^2(s,\Omega). \tag{A14}$$

We transform the coordinate system in the Eq. (A14) from a spherical coordinate system centered at the Earth's location to a spherical coordinate system centered at the dwarf galaxy considered. Let $\vec{r'}$ be the position from the center of the galaxy; the J factor in this coordinate system is given by

$$\begin{aligned} J &= \int \mathrm{d}s \int \mathrm{d}\Omega \rho^2(s,\Omega) = \int \mathrm{d}V \frac{\rho^2(s,\Omega)}{s^2} \\ &= \int \mathrm{d}V' \frac{\rho^2(r',\Omega')}{r'^2 - 2d_c r' \cos\theta' + d_c^2}, \end{aligned} \tag{A15}$$

where $d_c = |\vec{r}|$ is the comoving distance and $\vec{r}\cdot\vec{r'} = d_c r' \cos\theta'$. Note that the volume elements between the two coordinate systems are equal, $\mathrm{d}V = \mathrm{d}V'$, since the transformation is only a translation. We integrate Eq. (A15) up to the virial radius r_{vir}, defined as the radius within which the mean density of the halo is equal to a specified overdensity factor times the critical density of the Universe, $\overline{\rho}(r < r_{\mathrm{vir}}) = \Delta_{\mathrm{vir}}\rho_c$. As per common convention, we take $\Delta_{\mathrm{vir}} = 200$. We assume the distant-source approximation $d_c \gg r_{\mathrm{vir}} > r'$, justified for most Milky Way dwarf galaxies, which allows us to write J as a volume integral in spherical coordinates [85]

$$J = \int \mathrm{d}V' \frac{\rho^2(r',\Omega')}{r'^2 - 2d_c r' \cos\theta' + d_c^2} \approx \frac{1}{d_c^2} \int \mathrm{d}V' \rho^2(r'). \tag{A16}$$

Plugging the GNFW density profile Eq. (A3) into the above expression, we find

$$J = \frac{1}{d_c^2} \int \mathrm{d}V' (\rho_{\mathrm{DM}}^{\mathrm{GNFW}})^2(r') = \frac{4\pi\rho_s^2 r_s^2}{d_c^2} \int_0^{r_{\mathrm{vir}}} \mathrm{d}r' \left(\frac{r}{r_s}\right)^{2-2\gamma} \left(1 + \frac{r}{r_s}\right)^{2\gamma-6}, \tag{A17}$$

$$= \frac{4\pi\rho_s^2 r_s^3}{d_c^2} \left[\frac{c_{\mathrm{vir}}^{3-2\gamma}(1+c_{\mathrm{vir}})^{2\gamma-5}(20 + 4\gamma^2 + 2c_{\mathrm{vir}}(5+c_{\mathrm{vir}}) - 2\gamma(9+2c_{\mathrm{vir}}))}{(5-2\gamma)(4-2\gamma)(3-2\gamma)}\right], \tag{A18}$$

where $c_{\mathrm{vir}} \equiv r_{\mathrm{vir}}/r_s$ is the virial concentration of the dwarf galaxy. Note that the integration only converges when $\gamma < 1.5$, which is the case for the DM profiles shown in the main text.

4. Prior distributions on the parameters of interest

The prior distributions for both our method and the Jeans analysis are shown in Table I. The priors on the DM density

profile parameters are the same between the Jeans analysis and our method. As described in Sec. A 2, the Jeans analysis consists of an initial fit of the light profile parameters L and $r_\star$, and a subsequent joint fit of both the DM and light profile parameters.

APPENDIX B: ADDITIONAL RESULTS

1. Systematic variations on the analysis

We explore variations on the assumptions made in our baseline analysis, including the effect of varying the mean number of stars $\mu_{\rm stars}$ and line-of-sight velocity measurement error Δv. We also examine the performance of different graph convolution schemes.

a. Variations on the number of stars

We generate three datasets (including the one presented in the main text) with the same velocity measurement error $\Delta = 0.1$ km/s and different mean number of stars: $\mu_{\rm stars} = 20, 100, 1000$. As in the baseline case, each dataset contain 80 000 training samples, 10 000 validation samples, and 10 000 test samples. For each dataset, we train a GNN and normalizing flow with the same hyperparameters as described in the main text and plot the results on the test samples in Fig. 7. We take the marginal medians as the predicted DM parameters, bin them based on their truth values, and show the median (solid blue line), middle-68% (blue bands), and middle-95% (light blue bands) containment regions. The baseline $\mu_{\rm stars} = 100$ stars (middle column) case is the same as that in the main text and is shown here for comparison.

The performance on the scale radius r_s (middle row) and density normalization ρ_s (bottom row) is similar for all three cases. For the inner slope γ (top row), we see that increasing the number of stars helps increase the accuracy of the marginal medians. This is expected because the underlying phase-space distribution is more completely sampled for a larger number of observed tracer stars. We observed a similar increase in prediction accuracy with increasing sample size in Ref. [30]. The marginal medians for the $\mu_{\rm stars} = 20$ case are slightly biased towards the tails of the γ prior distribution. This is potentially due to clipping caused by a finite prior range that prevents the marginal median from being centered around the true value when constraining power is low.

b. Variations on measurement uncertainty

We generate three datasets with the same mean number of stars $\mu_{\rm stars}$ and different velocity measurement errors $\Delta v = 0.1, 1.0, 2.5$ km/s. Similar to our baseline case of $\Delta v = 0.1$ km/s, each dataset has 80 000 training samples, 10 000 validation samples, and 10 000 test samples, and we train our pipeline on each dataset separately. The results of this variation are shown in Fig. 8 in the same format as Fig. 7 (with the baseline $\Delta v = 0.1$ km/s case the same as in the result of the main text).

Again, we observe that for the scale radius r_s (middle row) and the density normalization ρ_0 (bottom row), the performance is approximately constant across all variations. Similar to Ref. [30], as the measurement uncertainties increase we observe a decrease in constraining power of the central slope γ. This is to be expected because we did not explicitly account for the uncertainties in the neural network architecture—sampling from the noise model is treated as a form of data augmentation, with larger error magnitudes leading to increased sample variance and reduced sample efficiency. We defer the explicit inclusion of observational uncertainties in the neural network construction to future work.

c. Variations on the graph convolution scheme

In Fig. 9 we show variations on the type of graph convolutional layer used, otherwise keeping all hyperparameters (e.g., channel dimension) the same as the baseline case. This case, labeled "ChebConv" and shown in the left-most column, uses the graph convolution prescription from Ref. [53].

The second column shows results using the graph attention layer from Ref. [86], which uses the attention mechanism to implicitly weigh neighboring nodes. The final column shows results using the deep set architecture, where node-wise features are obtained using a dense network and aggregated through averaging, before finally being passed through a dense layer as in the baseline case. Relatively good recovery of all parameters can be seen in these cases. Together, these results point to the fact that aggregation of neighborhood information may not play a key role in the success of our method on the examples tested. We expect this fact to not hold on more realistic test cases—in particular, the systems in this paper were chosen to be relatively simple (spherical and dynamically equilibrated) in order to enable a direct comparison with the conventional Jeans analysis method. With these assumptions and our choice of stellar and DM profiles, the node-wise features $\{R, v_{\rm los}\}$ can be assumed to be independent.

Finally, the third column of Fig. 9 shows results using the graph convolutional layer from Ref. [87] with the default configuration in PyTorch Geometric, showing poor recovery of the inner slope γ.

2. Comparison of inner density slopes obtained using Jeans analysis and GNN + Flow

In order to quantitatively compare the inferred inner density slope γ posteriors for a large galaxy sample, we compute and summarize the Jensen-Shannon (JS) divergence [88,89] between 100 pairs of cored and cuspy γ posteriors using our method as well as Jeans analysis. Given two probability distribution functions $P(x)$ and

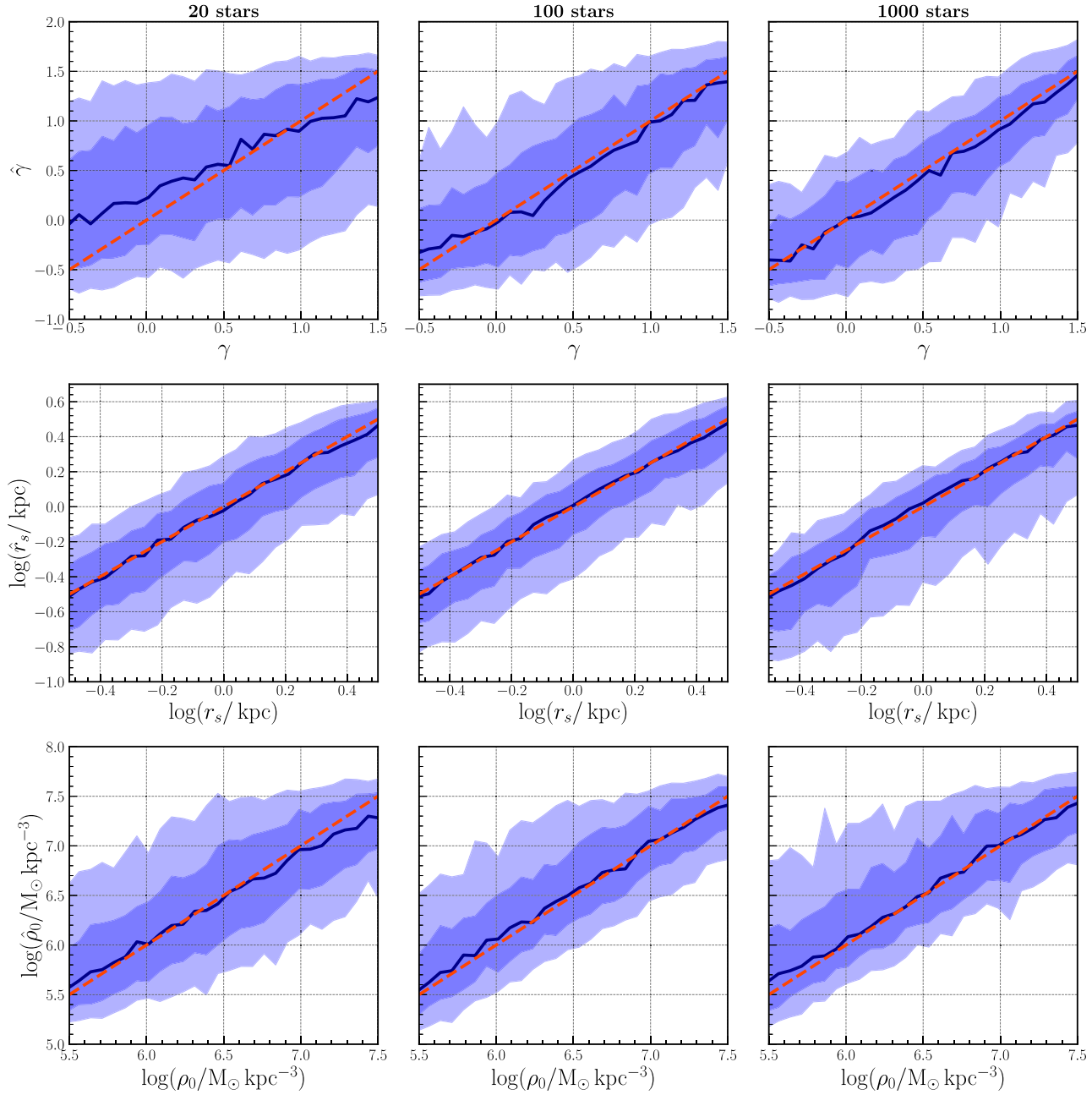

FIG. 7. The predicted DM parameters versus truth DM parameters for three different runs. In each run, a GNN and normalizing is trained and tested on galaxies with a mean number of stars $\mu_{\mathrm{stars}} = 20$ (left), 100 (center), 1000 (right) stars. The line-of-sight velocity measurement error is set to be $\Delta v = 0.1$ km/s. The $\mu_{\mathrm{stars}} = 100$ case (middle column) is the same case as shown in the result of the main text. The median (solid blue line), middle-68% percentile (blue band), and middle-95% (light blue band) containment regions of each bin are shown. The diagonal dashed red line shows the locus of equality between true and predicted values.

$Q(x)$, defined over the same probability space $\mathcal{X} \ni x$, the JS divergence is defined as

$$D_{\mathrm{JS}}(P\|Q) \equiv \frac{1}{2}[D_{\mathrm{KL}}(P\|M) + D_{\mathrm{KL}}(Q\|M)], \quad \text{(B1)}$$

where $M(x) = P(x) + Q(x)$ and D_{KL} is the Kullback-Leibler (KL) divergence, for which we use the definition

$$D_{\mathrm{KL}}(P\|Q) = \sum_{x\in\mathcal{X}} P(x)\log_2\left(\frac{P(x)}{Q(x)}\right). \quad \text{(B2)}$$

The KL divergence of P from Q represents the expected entropy gain from using Q as an approximation for truth distribution P [90]. The JS divergence is based on the KL divergence but has more desirable properties for the present case, specifically being symmetric in the two posterior

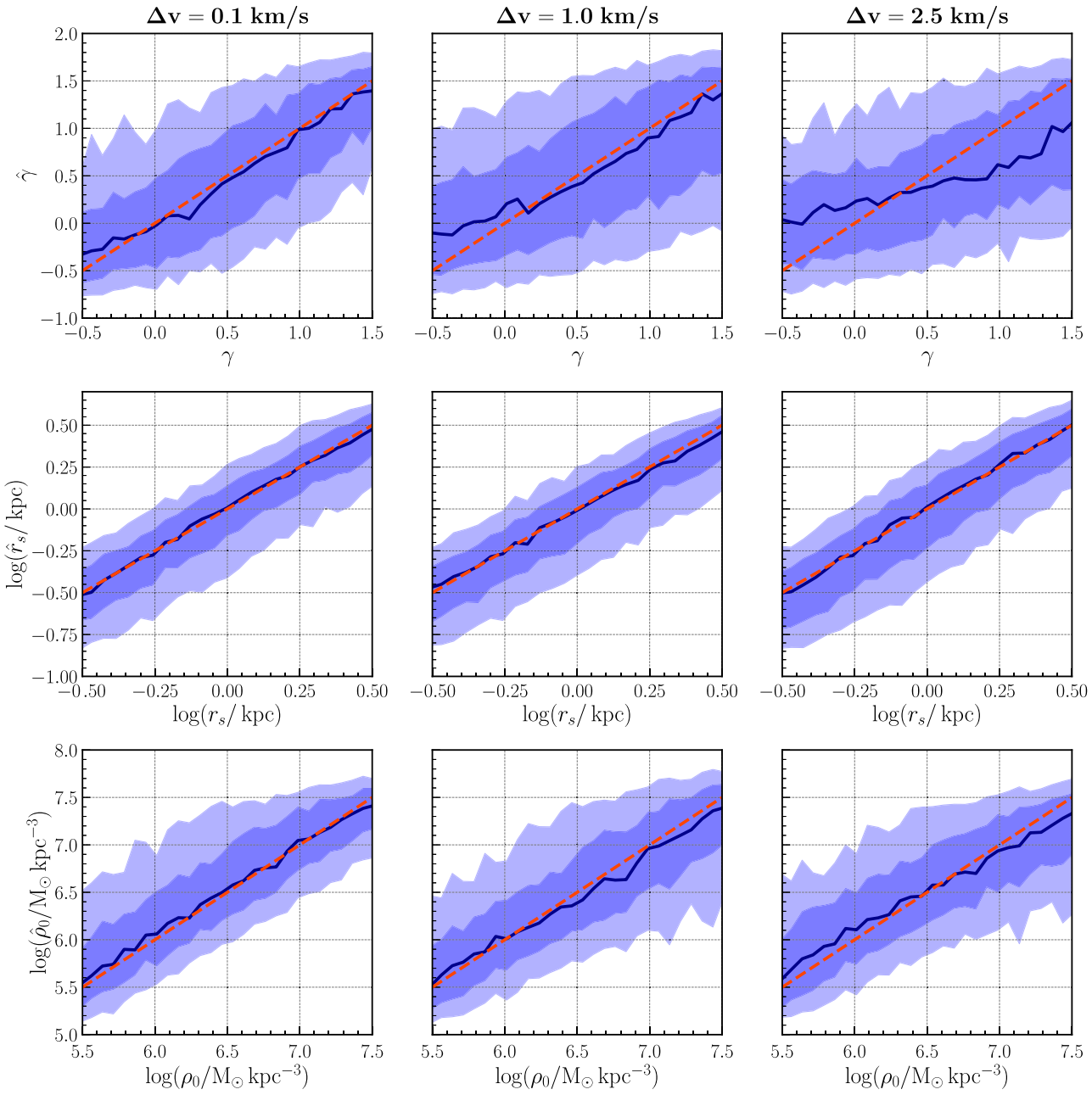

FIG. 8. Same as Fig. 7, but with different line-of-sight velocity errors Δv. The mean number of stars μ_{stars} is 100 stars. The $\Delta v = 0.1$ km/s case (left column) is the same case as shown in the result of the main text.

distributions. Note that using a $\log_2$ definition of the constituent KL divergences, the JS divergence is constrained to lie within the range $D_{\mathrm{JS}} \in [0, 1]$, with 0 corresponding to two identical distributions and 1 two nonoverlapping distributions.

We generate 100 galaxies with cored DM profiles ($\gamma = 0$) and 100 galaxies with cuspy profiles ($\gamma = 1$). These galaxies have the same DM scale radius $r_s = 1$ kpc, density normalization $\rho_0 = 10^7$ $\mathrm{M}_\odot/\mathrm{kpc}^3$, and light profile parameters as the two galaxies presented in the result of the main text. We obtain samples the DM parameter posteriors for each galaxy and calculate D_{JS} using Eq. (B1) for each pair of cored and cuspy galaxies (in total, there are 10 000 pairs). We show the distribution of the JS divergences D_{JS} between the $\gamma = 0$ posterior and $\gamma = 1$ posterior for the Jeans analysis and for our method in Fig. 10. The median and middle-68% containment region values are $D_{\mathrm{JS}}^{\mathrm{Jeans}} = 0.417^{+0.288}_{-0.280}$ and $D_{\mathrm{JS}}^{\mathrm{GNN}} = 0.629^{+0.196}_{-0.278}$. Our method generically produces higher values of D_{JS} compared to the Jeans analysis, corresponding to a larger contrast between the cored and cuspy γ posteriors and thus increased ability to distinguish between the two scenarios given a set of observations.

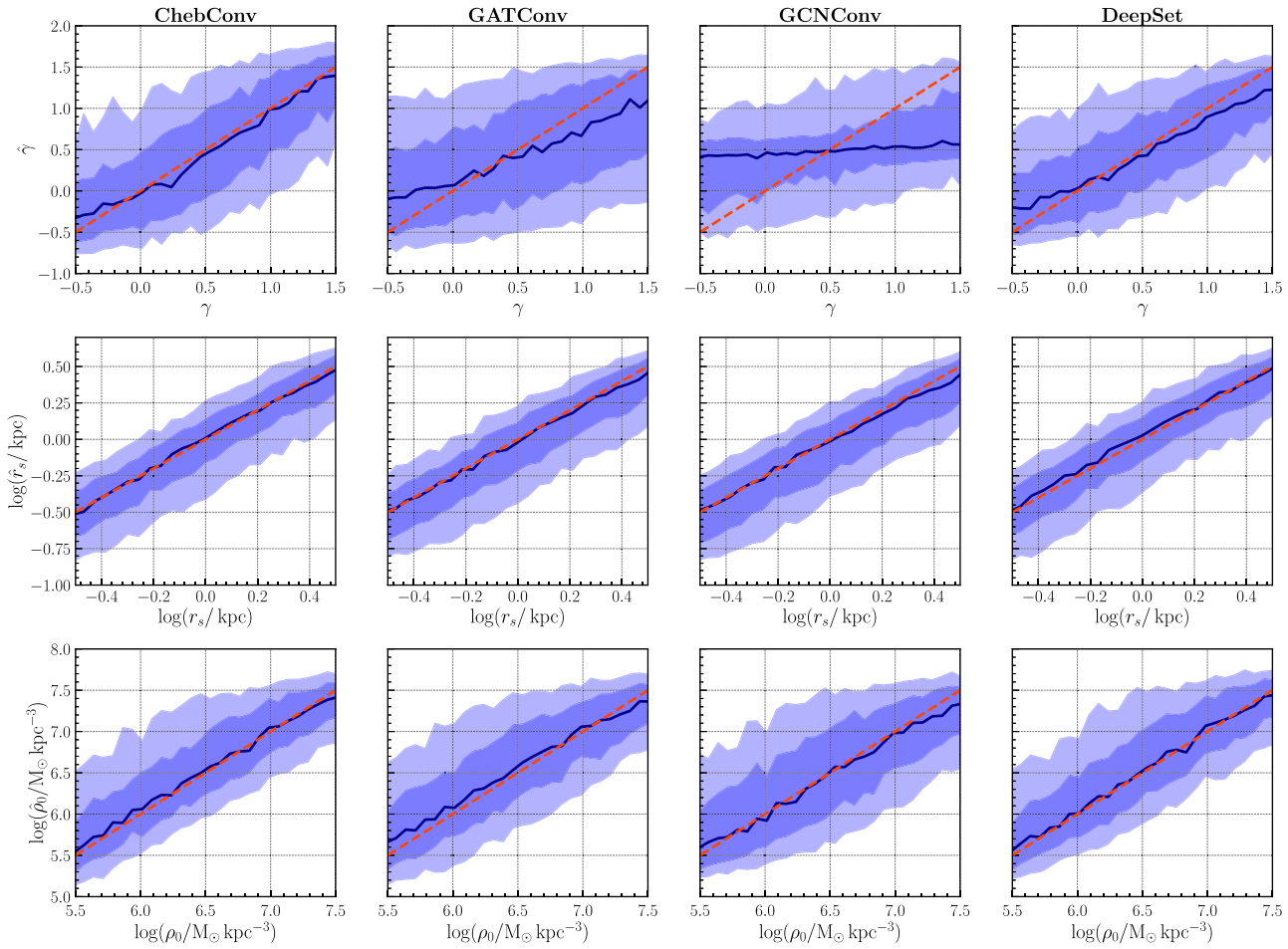

FIG. 9. Same as Fig. 7, but with variation in the type of graph convolutional layer. Each network is trained and tested on the dataset presented in the main text, i.e., $\mu_{\mathrm{stars}} = 100$ stars and $\Delta v = 0.1$ km/s. The ChebConv case (left column) refers to the baseline graph-convolutional scheme presented in the main text.

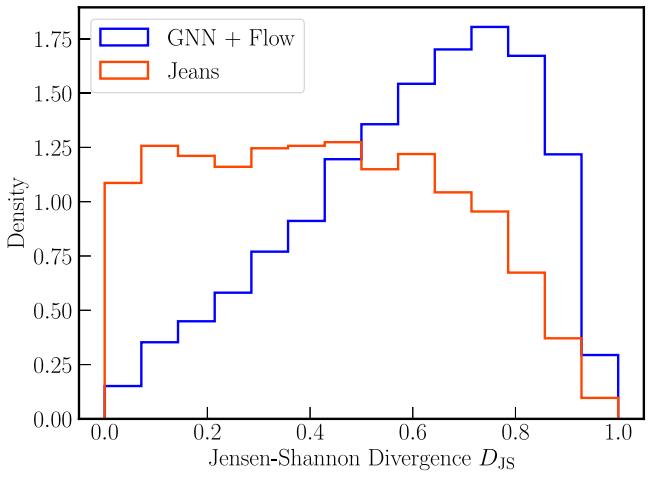

FIG. 10. The JS divergence D_{JS} between the $\gamma = 0$ and $\gamma = 1$ posteriors for the Jeans analysis (red) and the GNN and normalizing flow model (blue). As evidenced from the higher values of the D_{JS}, compared to the Jeans analysis, the GNN and normalizing flow produces $\gamma = 0$ and $\gamma = 1$ more distinct from each other.

3. Comparison of J factors inferred using Jeans analysis and GNN + Flow

As mentioned in the main text, we calculate the J factors (normalized to distance of 100 kpc) for 100 galaxies randomly sampled from our test set. In this section, we plot the posterior distribution of the J factor for a few examples for the Jeans analysis and our method in Fig. 11. In all cases shown, the GNN and normalizing flow can predict the J factor with a similar or higher accuracy compared to the Jeans analysis.

4. Test of statistical coverage of the inferred posteriors

Simulation-based inference methods such as those employed in this work can be susceptible to producing overconfident posteriors [91]. In this section, we examine the quality of the posterior distributions produced by our simulation-based inference pipeline following the prescription in Ref. [91].

Using the same notation as in the main text, let θ be the parameters of interest (i.e., the DM and stellar parameters)

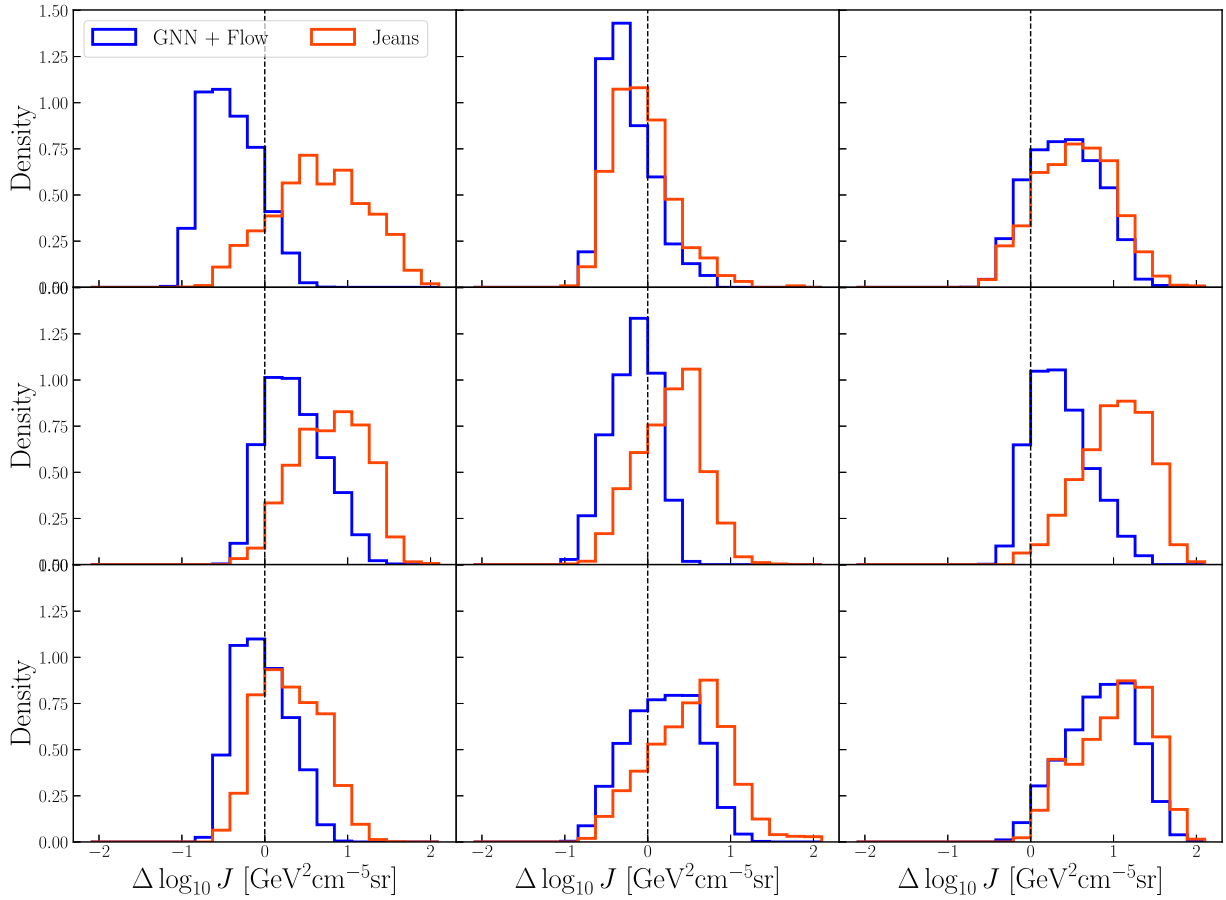

FIG. 11. Posterior distributions of the J factors predicted by the Jeans analysis (red) and by the GNN and normalizing flow (blue). Each panel show the $\Delta \log_{10} J = \log_{10} J_{\text{predict}} - \log_{10} J_{\text{truth}}$ posteriors of an example galaxy. The dashed black line represents the truth J-factor value (i.e., $\Delta \log_{10} J = 0$).

and x be the observable (i.e., the graph representation of a galaxy constructed from its the stellar kinematics). We denote our learned posterior density estimator as $\hat{p}(\theta|x)$. For a confidence level $1-\alpha$, the expected coverage probability is

$$\mathbb{E}_{(\theta,x)\sim p(\theta,x)}[\mathbb{1}_\Theta(\theta \in \Theta_{\hat{p}(\theta|x)}(1-\alpha))], \tag{B3}$$

where $\Theta_{\hat{p}(\theta|x)}(1-\alpha)$ gives the $1-\alpha$ highest posterior density interval of the estimator $\hat{p}(\theta|x)$ and $\mathbb{1}_\Theta()$ is an indicator function mapping samples that fall within the highest posterior density interval to unity. Given N samples from the joint distribution $(\theta^\star, x) \sim p(\theta, x)$, the empirical expected coverage for the posterior estimator $\hat{p}(\theta|x)$ is defined as

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_\Theta(\theta^\star \in \Theta_{\hat{p}(\theta|x)}(1-\alpha)). \tag{B4}$$

The nominal expected coverage is the expected coverage in the case when $\hat{p}(\theta|x) = p(\theta|x)$ and equals to the confidence level $1-\alpha$. In general, we want our estimator to have an empirical expected coverage probability larger than or equal to the nominal expected coverage probability at all confidence levels. Such estimators will produce conservative posteriors, in contrast to overconfident posteriors which may spuriously exclude allowable regions of parameter space.

In Fig. 12, we plot the empirical expected coverage probability for the marginal posteriors as produced by the baseline model presented in the main text (i.e., $\mu_{\text{stars}} = 100$ and $\Delta v = 0.1$ km/s) against the nominal confidence levels. The dashed diagonal black line represents the nominal expected coverage probability. A conservative estimator will lie above the diagonal line, while an overconfident estimator will lie below it. In general, the posteriors produced by our model lie very close to the well-calibrated regime, although the posteriors for the stellar radius $r_\star$ are slightly overconfident. We note that methodological improvement in calibration quality of posteriors produced using forward-modeling approaches is an ongoing, active area of research [92].

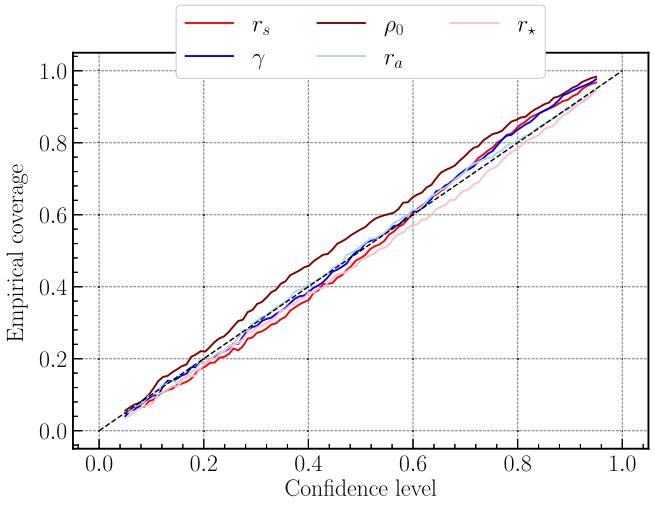

FIG. 12. The expected coverage of the marginal DM and stellar parameters posteriors by the model presented in the main text ($\mu_{\text{stars}} = 100$, $\Delta v = 0.1$ km/s). If an estimator produces perfectly calibrated posteriors, then its empirical expected coverage probability is equal to the nominal expected coverage probability (dashed diagonal black line). The estimator is conservative (overconfident) if it produces an empirical expected coverage probability above (below) the diagonal line.

5. Robustness to observational projection

Typically, only line-of-sight velocities and angular coordinates of tracer stars are observationally accessible, presenting the challenge of working with incomplete phase-space information when inferring the DM density profile. A test of our method is then its susceptibility to the specific direction from which a dwarf galaxy is viewed—its observational projection. Since different projections correspond to the same latent DM parameters, it is desirable for different projections to produce similar summary representations, and therefore similar posterior distributions. Since this is not explicitly baked into the network, approximate projective symmetry can be learned implicitly using the training sample.

We take the galaxies with the same DM profiles presented in the main text and project them onto orthogonal planes. For each galaxy, we thus obtain three orthogonal projections (including the original projection). In Fig. 13, we show the 68% contour line of the posterior DM parameters for three orthogonal projections of two galaxies (one with a cored DM profile and one with a cuspy DM profile) and the graph representations of the projections. For each projection, its graph representation matches the

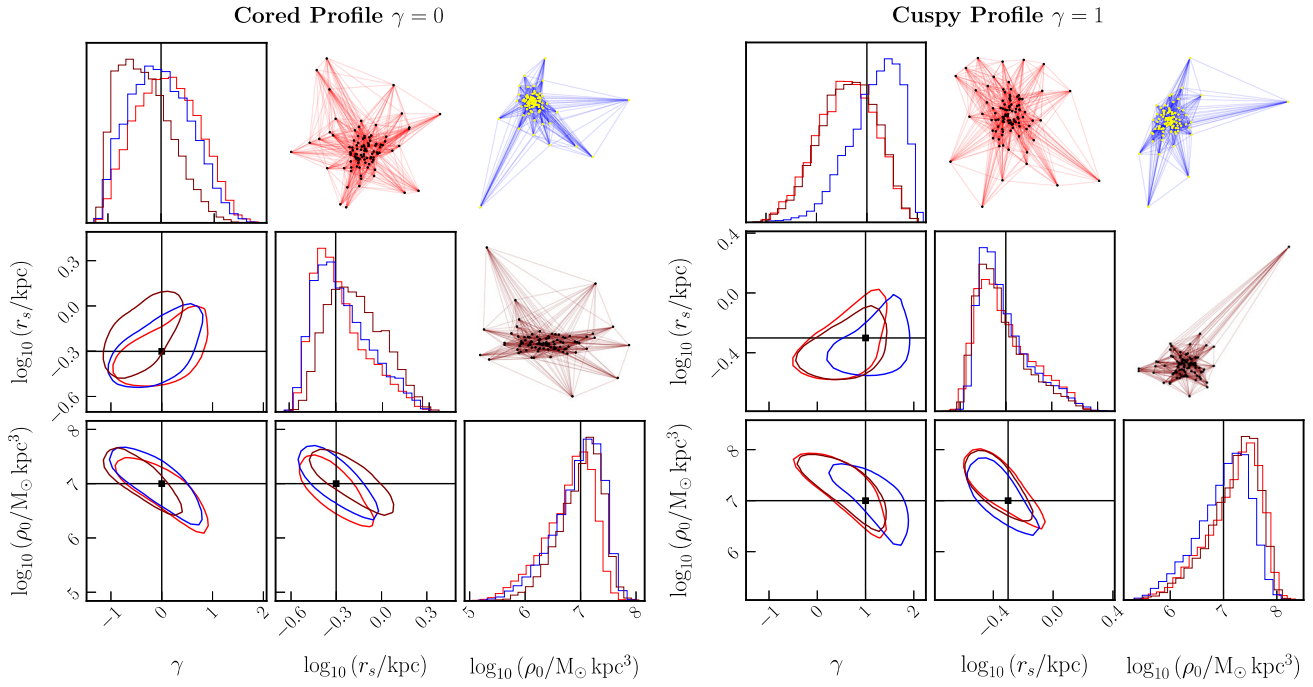

FIG. 13. Example corner plots of the posterior DM parameters of two test galaxies, each with three orthogonal projections. The left (right) panel shows the posteriors for a galaxy with a cored (cuspy) DM profile. The contour lines show the 68% containment region, with each color representing a different projection. For each projection, its graph representation is also shown (with the edge color matching the contour color) on the upper right corner.

color of the contour line of the DM posteriors. It can be seen that even though the graph representations may vary significantly between projections (e.g., the positions of each node may shift, forming new edge connections or breaking up old ones), the DM posteriors remain consistent. Note that we do not expect the DM posteriors to be identical, since the information given to the GNN is not the identical between projections.

[1] S. D. M. White and M. J. Rees, Core condensation in heavy halos: A two-stage theory for galaxy formation and clustering., Mon. Not. R. Astron. Soc. **183**, 341 (1978).

[2] C. Pryor and J. Kormendy, The dark matter halos of Draco and Ursa minor, Astron. J. **100**, 127 (1990).

[3] M. Cirelli, G. Corcella, A. Hektor, G. Hütsi, M. Kadastik, P. Panci, M. Raidal, F. Sala, and A. Strumia, PPPC 4 DM ID: A poor particle physicist cookbook for dark matter indirect detection, J. Cosmol. Astropart. Phys. 03 (2011) 051.

[4] M. Ackermann *et al.* (The Fermi-LAT Collaboration), Searching for Dark Matter Annihilation from Milky Way Dwarf Spheroidal Galaxies with Six Years of Fermi Large Area Telescope Data, Phys. Rev. Lett. **115**, 231301 (2015).

[5] G. D. Martinez, A robust determination of Milky Way satellite properties using hierarchical mass modelling, Mon. Not. R. Astron. Soc. **451**, 2524 (2015).

[6] A. Geringer-Sameth, S. M. Koushiappas, and M. Walker, Dwarf galaxy annihilation and decay emission profiles for dark matter experiments, Astrophys. J. **801**, 74 (2015).

[7] A. Geringer-Sameth and S. M. Koushiappas, Exclusion of Canonical Weakly Interacting Massive Particles by Joint Analysis of Milky Way Dwarf Galaxies with Data from the Fermi Gamma-Ray Space Telescope, Phys. Rev. Lett. **107**, 241303 (2011).

[8] M. Ackermann *et al.* (The Fermi-LAT Collaboration), Constraining Dark Matter Models from a Combined Analysis of Milky Way Satellites with the Fermi Large Area Telescope, Phys. Rev. Lett. **107**, 241302 (2011).

[9] M. Mazziotta, F. Loparco, F. de Palma, and N. Giglietto, A model-independent analysis of the Fermi Large Area Telescope gamma-ray data from the Milky Way dwarf galaxies and halo to constrain dark matter scenarios, Astropart. Phys. **37**, 26 (2012).

[10] A. Geringer-Sameth, S. M. Koushiappas, and M. G. Walker, Comprehensive search for dark matter annihilation in dwarf galaxies, Phys. Rev. D **91**, 083535 (2015).

[11] M. Ackermann *et al.* (Fermi-LAT Collaboration), Dark matter constraints from observations of 25 Milky Way satellite galaxies with the Fermi Large Area Telescope, Phys. Rev. D **89**, 042001 (2014).

[12] B. Anderson, J. Chiang, J. Cohen-Tanugi, J. Conrad, A. Drlica-Wagner, M. L. Garde, and S. Zimmer, Using likelihood for combined data set analysis (2015), https://inspirehep.net/literature/1343772.

[13] J. F. Navarro, V. R. Eke, and C. S. Frenk, The cores of dwarf galaxy haloes, Mon. Not. R. Astron. Soc. **283**, L72 (1996).

[14] K. Spekkens, R. Giovanelli, and M. P. Haynes, The Cusp/Core problem in galactic halos: Long-slit spectra for a large dwarf galaxy sample, Astron. J. **129**, 2119 (2005).

[15] J. F. Navarro, C. S. Frenk, and S. D. M. White, A universal density profile from hierarchical clustering, Astrophys. J. **490**, 493 (1997).

[16] S.-H. Oh, C. Brook, F. Governato, E. Brinks, L. Mayer, W. J. G. de Blok, A. Brooks, and F. Walter, The central slope of dark matter cores in dwarf galaxies: Simulations versus THINGS, Astron. J. **142**, 24 (2011).

[17] S.-H. Oh, D. A. Hunter, E. Brinks, B. G. Elmegreen, A. Schruba, F. Walter, M. P. Rupen, L. M. Young, C. E. Simpson, M. C. Johnson, K. A. Herrmann, D. Ficut-Vicas, P. Cigan, V. Heesen, T. Ashley, and H.-X. Zhang, High-resolution mass models of dwarf galaxies from LITTLE THINGS, Astron. J. **149**, 180 (2015).

[18] J. S. Bullock and M. Boylan-Kolchin, Small-Scale challenges to the ΛCDM paradigm, Annu. Rev. Astron. Astrophys. **55**, 343 (2017).

[19] J. F. Navarro, V. R. Eke, and C. S. Frenk, The cores of dwarf galaxy haloes, Mon. Not. R. Astron. Soc. **283**, L72 (1996).

[20] J. I. Read and G. Gilmore, Mass loss from dwarf spheroidal galaxies: The origins of shallow dark matter cores and exponential surface brightness profiles, Mon. Not. R. Astron. Soc. **356**, 107 (2005).

[21] S. Mashchenko, H. M. P. Couchman, and J. Wadsley, The removal of cusps from galaxy centres by stellar feedback in the early Universe, Nature (London) **442**, 539 (2006).

[22] A. Pontzen and F. Governato, How supernova feedback turns dark matter cusps into cores, Mon. Not. R. Astron. Soc. **421**, 3464 (2012).

[23] M. R. Lovell, V. Eke, C. S. Frenk, L. Gao, A. Jenkins, T. Theuns, J. Wang, S. D. M. White, A. Boyarsky, and O. Ruchayskiy, The haloes of bright satellite galaxies in a warm dark matter universe, Mon. Not. R. Astron. Soc. **420**, 2318 (2012).

[24] O. D. Elbert, J. S. Bullock, S. Garrison-Kimmel, M. Rocha, J. Oñorbe, and A. H. G. Peter, Core formation in dwarf haloes with self-interacting dark matter: No fine-tuning necessary, Mon. Not. R. Astron. Soc. **453**, 29 (2015).

[25] J. D. Burger, J. Zavala, L. V. Sales, M. Vogelsberger, F. Marinacci, and P. Torrey, Degeneracies between self-interacting dark matter and supernova feedback as cusp-core transformation mechanisms, Mon. Not. R. Astron. Soc. **513**, 3458 (2022).

[26] J. H. Jeans, On the theory of star-streaming and the structure of the universe, Mon. Not. R. Astron. Soc. **76**, 70 (1915).

[27] V. Bonnivard, C. Combet, D. Maurin, and M. G. Walker, Spherical Jeans analysis for dark matter indirect detection in dwarf spheroidal galaxies—impact of physical parameters and triaxiality, Mon. Not. R. Astron. Soc. **446**, 3002 (2015).

[28] K. El-Badry, A. R. Wetzel, M. Geha, E. Quataert, P. F. Hopkins, D. Kereš, T. K. Chan, and C.-A. Faucher-Giguère, When the jeans do not fit: How stellar feedback drives stellar kinematics and complicates dynamical modeling in low-mass galaxies, Astrophys. J. **835**, 193 (2017).
[29] A. Genina, J. I. Read, C. S. Frenk, S. Cole, A. Benítez-Llambay, A. D. Ludlow, J. F. Navarro, K. A. Oman, and A. Robertson, To β or not to β: Can higher order Jeans analysis break the mass–anisotropy degeneracy in simulated dwarfs?, Mon. Not. R. Astron. Soc. **498**, 144 (2020).
[30] L. J. Chang and L. Necib, Dark matter density profiles in dwarf galaxies: Linking Jeans modelling systematics and observation, Mon. Not. R. Astron. Soc. **507**, 4715 (2021).
[31] A. Helmi, Streams, substructures, and the early history of the Milky Way, Annu. Rev. Astron. Astrophys. **58**, 205 (2020).
[32] J. I. Read, G. A. Mamon, E. Vasiliev, L. L. Watkins, M. G. Walker, J. Peñarrubia, M. Wilkinson, W. Dehnen, and P. Das, Breaking beta: A comparison of mass modelling methods for spherical systems, Mon. Not. R. Astron. Soc. **501**, 978 (2021).
[33] J. I. Read and P. Steger, How to break the density-anisotropy degeneracy in spherical stellar systems, Mon. Not. R. Astron. Soc. **471**, 4541 (2017).
[34] J. Read, M. Walker, and P. Steger, The case for a cold dark matter cusp in Draco, Mon. Not. R. Astron. Soc. **481**, 860 (2018).
[35] M. G. Walker and J. Peñarrubia, A method for measuring (Slopes of) the mass profiles of dwarf spheroidal galaxies, Astrophys. J. **742**, 20 (2011).
[36] N. C. Amorisco and N. W. Evans, Dark matter cores and cusps: The case of multiple stellar populations in dwarf spheroidals, Mon. Not. R. Astron. Soc. **419**, 184 (2012).
[37] L. Zhu, G. van de Ven, L. L. Watkins, and L. Posti, A discrete chemo-dynamical model of the dwarf spheroidal galaxy Sculptor: Mass profile, velocity anisotropy and internal rotation, Mon. Not. R. Astron. Soc. **463**, 1117 (2016).
[38] L. E. Strigari, J. S. Bullock, and M. Kaplinghat, Determining the nature of dark matter with astrometry, Astrophys. J. Lett. **657**, L1 (2007).
[39] L. E. Strigari, C. S. Frenk, and S. D. White, Dynamical constraints on the dark matter distribution of the Sculptor dwarf spheroidal from stellar proper motions, Astrophys. J. **860**, 56 (2018).
[40] G. A. Mamon, A. Biviano, and G. Boué, MAMPOSSt: Modelling anisotropy and mass profiles of observed spherical systems—I. Gaussian 3D velocities, Mon. Not. R. Astron. Soc. **429**, 3079 (2013).
[41] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis*, Chapman & Hall/CRC Texts in Statistical Science (Chapman & Hall/CRC, London, 2003).
[42] A. F. M. Smith and A. E. Gelfand, Bayesian statistics without tears: A sampling–resampling perspective, Am. Stat. **46**, 84 (1992).
[43] D. B. Rubin, Using the SIR Algorithm to Simulate Posterior Distributions, in *Bayesian Statistics 3: Proceedings of the Third Valencia International Meeting*, edited by J. M. Bernardo, M. H. Degroot, D. V. Lindley, and A. M. Smith (Clarendon Press, Oxford, 1988), pp. 385–402.
[44] H. C. Plummer, On the problem of distribution in globular star clusters, Mon. Not. R. Astron. Soc. **71**, 460 (1911).
[45] L. P. Osipkov, Spherical systems of gravitating bodies with an ellipsoidal velocity distribution, Pisma Astron. Zh. **5**, 77 (1979).
[46] D. Merritt, Spherical stellar systems with spheroidal velocity distributions, Astron. J. **90**, 1027 (1985).
[47] J. D. Simon and M. Geha, The kinematics of the ultra-faint Milky Way satellites: Solving the missing satellite problem, Astrophys. J. **670**, 313 (2007).
[48] M. Mateo, E. W. Olszewski, and M. G. Walker, The velocity dispersion profile of the remote dwarf spheroidal galaxy Leo I: A tidal hit and run?, Astrophys. J. **675**, 201 (2008).
[49] M. G. Walker, M. Mateo, and E. W. Olszewski, Stellar velocities in the Carina, Fornax, Sculptor, and Sextans dSph galaxies: Data from the Magellan/MMFS survey, Astrophys. J. **137**, 3100 (2009).
[50] M. G. Walker, M. Mateo, E. W. Olszewski, J. Peñarrubia, N. W. Evans, and G. Gilmore, A universal mass profile for dwarf spheroidal galaxies?, Astrophys. J. **704**, 1274 (2009).
[51] T. L. Makinen, T. Charnock, P. Lemos, N. Porqueres, A. Heavens, and B. D. Wandelt, The cosmic graph: Optimal information extraction from large-scale structure using catalogues, arXiv:2207.05202.
[52] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, Graph attention networks, arXiv:1710.10903.
[53] M. Defferrard, X. Bresson, and P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, arXiv:1606.09375.
[54] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, Normalizing flows for probabilistic modeling and inference, J. Mach. Learn. Res. **22**, 1 (2021).
[55] D. J. Rezende and S. Mohamed, Variational inference with normalizing flows, in *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 2015*, JMLR Workshop and Conference Proceedings, Vol. 37, edited by F. R. Bach and D. M. Blei (2015), pp. 1530–1538.
[56] G. Papamakarios, T. Pavlakou, and I. Murray, Masked autoregressive flow for density estimation, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17 (Curran Associates Inc., Red Hook, NY, USA, 2017), pp. 2335–2344.
[57] M. Germain, K. Gregor, I. Murray, and H. Larochelle, MADE: Masked autoencoder for distribution estimation, in *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 2015*, JMLR Workshop and Conference Proceedings, Vol. 37, edited by F. R. Bach and D. M. Blei (2015), pp. 881–889.
[58] K. Cranmer, J. Brehmer, and G. Louppe, The frontier of simulation-based inference, Proc. Natl. Acad. Sci. U.S.A. **117**, 30055 (2020).
[59] K. Cranmer and G. Louppe, Unifying generative models and exact likelihood-free inference with conditional bijections, J. Brief Ideas (2016).
[60] G. Papamakarios and I. Murray, Fast ϵ-free inference of simulation models with Bayesian conditional density estimation, in *Proceedings of the 30th International Conference*

on Neural Information Processing Systems, NIPS'16 (Curran Associates Inc., Red Hook, NY, USA, 2016), pp. 1036–1044.

[61] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.

[62] I. Loshchilov and F. Hutter, Decoupled weight decay regularization, in *Proceedings of the International Conference on Learning Representations* (2019), https://openreview.net/forum?id=Bkg6RiCqY7.

[63] J. Skilling, Nested sampling, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, American Institute of Physics Conference Series, Vol. 735, edited by R. Fischer, R. Preuss, and U. V. Toussaint (2004), pp. 395–405, 10.1063/1.1835238.

[64] J. Skilling, Nested sampling for general Bayesian computation, Bayesian Anal. **1**, 833 (2006).

[65] J. S. Speagle, DYNESTY: A dynamic nested sampling package for estimating Bayesian posteriors and evidences, Mon. Not. R. Astron. Soc. **493**, 3132 (2020).

[66] S. Garrison-Kimmel, P. F. Hopkins, A. Wetzel, J. S. Bullock, M. Boylan-Kolchin, D. Kereš, C.-A. Faucher-Giguère, K. El-Badry, A. Lamberts, E. Quataert, and R. Sanderson, The Local Group on FIRE: Dwarf galaxy populations across a suite of hydrodynamic simulations, Mon. Not. R. Astron. Soc. **487**, 1380 (2019).

[67] T. Nguyen, S. Mishra-Sharma, R. Williams, L. Necib, trivnguyen/dsph_gnn: Uncovering dark matter density profiles in dwarf galaxies with graph neural networks, v1.0.0 (2023), 10.5281/zenodo.7545540.

[68] http://iaifi.org/.

[69] S. T. Brown, P. Buitrago, E. Hanna, S. Sanielevici, R. Scibek, and N. A. Nystrom, Bridges-2: A platform for rapidly-evolving and data intensive research, in *Practice and Experience in Advanced Research Computing*, PEARC '21 (Association for Computing Machinery, New York, NY, USA, 2021).

[70] J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. Scott, and N. Wilkins-Diehr, XSEDE: Accelerating scientific discovery, Comput. Sci. Eng. **16**, 62 (2014).

[71] G. Ashton *et al.*, BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy, Astrophys. J. Suppl. Ser. **241**, 27 (2019).

[72] F. Perez and B. E. Granger, IPython: A system for interactive scientific computing, Comput. Sci. Eng. **9**, 21 (2007).

[73] T. Kluyver *et al.*, Jupyter notebooks—a publishing format for reproducible computational workflows, in *ELPUB* (2016), 10.3233/978-1-61499-649-1-87.

[74] J. D. Hunter, MATPLOTLIB: A 2D graphics environment, Comput. Sci. Eng. **9**, 90 (2007).

[75] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, nflows: Normalizing flows in PyTorch (2020).

[76] C. R. Harris, K. J. Millman, S. J. Van Der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith *et al.*, Array programming with NumPy, Nature (London) **585**, 357 (2020).

[77] A. Paszke *et al.*, PyTorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019), pp. 8024–8035, 10.5555/3454287.3455008.

[78] M. Fey and J. E. Lenssen, Fast graph representation learning with PyTorch Geometric, in *Proceedings of the ICLR Workshop on Representation Learning on Graphs and Manifolds* (2019), arXiv:1903.02428.

[79] W. Falcon *et al.*, Pytorchlightning/pytorch-lightning: 0.7.6 release (2020), 10.5281/zenodo.3828935.

[80] P. Virtanen *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in PYTHON, Nat. Methods **17**, 261 (2020).

[81] J. Binney and G. A. Mamon, M/L and velocity anisotropy from observations of spherical galaxies, or must M87 have a massive black hole, Mon. Not. R. Astron. Soc. **200**, 361 (1982).

[82] J. Binney and S. Tremaine, *Galactic Dynamics*, 2nd ed., Princeton Series in Astrophysics (Princeton University Press, Princeton, NJ, 2008).

[83] R. Barlow, Asymmetric statistical errors, in *Statistical Problems in Particle Physics, Astrophysics and Cosmology* (2004), pp. 56–59, 10.1142/p446.

[84] L. E. Strigari, S. M. Koushiappas, J. S. Bullock, M. Kaplinghat, J. D. Simon, M. Geha, and B. Willman, The most dark-matter-dominated galaxies: Predicted gamma-ray signals from the faintest Milky Way dwarfs, Astrophys. J. **678**, 614 (2008).

[85] M. Lisanti, S. Mishra-Sharma, N. L. Rodd, B. R. Safdi, and R. H. Wechsler, Mapping extragalactic dark matter annihilation with galaxy surveys: A systematic study of stacked group searches, Phys. Rev. D **97**, 063005 (2018).

[86] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, Graph attention networks, arXiv:1710.10903.

[87] T. N. Kipf and M. Welling, Semi-supervised classification with graph convolutional networks, arXiv:1609.02907.

[88] F. Nielsen, On the Jensen–Shannon symmetrization of distances relying on abstract means, Entropy **21**, 485 (2019).

[89] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing* (The MIT Press, Cambridge, USA, 1999).

[90] S. Kullback and R. A. Leibler, On information and sufficiency, Ann. Math. Stat. **22**, 79 (1951).

[91] J. Hermans, A. Delaunoy, F. Rozet, A. Wehenkel, and G. Louppe, A Trust Crisis In Simulation-Based Inference? Your Posterior Approximations Can Be Unfaithful (2021).

[92] B. Dey, D. Zhao, J. A. Newman, B. H. Andrews, R. Izbicki, and A. B. Lee, Calibrated predictive distributions via diagnostics for conditional coverage, arXiv:2205.14568.