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

Tri Nguyen,^{1,2} Siddharth Mishra-Sharma,^{1,2,3} Lina Necib^{1,2}

Abstract

Dwarf galaxies are small dark matter-dominated galaxies, some of which are embedded within the Milky Way; their lack of baryonic matter (stars and gas) makes them perfect test beds for dark matter detection. Understanding the distribution of dark matter in these systems can be used to pin down microphysical dark matter interactions that influence the formation and evolution of structures in our Universe. We introduce a new approach for inferring the dark matter density profiles of dwarf galaxies from the observable kinematics of stars bound to these systems using graph-based machine learning. Our approach aims to address some of the limitations of established methods based on dynamical Jeans modeling such as the necessity of assuming equilibrium and reliance on second-order moments of the stellar velocity distribution. We show that by leveraging more information about the available phase space of bound stars, this method can place stronger constraints on dark matter profiles in dwarf galaxies and has the potential to resolve some of the ongoing puzzles associated with the small-scale structure of dark matter halos.

1. Introduction

Cosmological structure formation is known to proceed hierarchically—smaller structures seed the formation of larger structures (White & Rees, 1978). 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 na-

ture of its self-interactions. Deviations from the canonical Λ Cold Dark Matter (Λ CDM) paradigm of cosmology would be imprinted in 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 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, making them versatile astrophysical laboratories for DM studies. A major goal in cosmology and particle physics is to detect non-gravitational interactions of DM. One of the main avenues to do so is DM indirect detection; DM would annihilate or decay into standard model (SM) particles, such as photons or electrons, which can be subsequently detected. Dwarf galaxies act as ideal targets for indirect detection searches as they are deficient in stars and gas, which would otherwise act as backgrounds for such studies.

An ongoing puzzle in cosmology is the so-called cusp-core problem, referring to whether the inner density profile of DM in dwarf galaxies is a cusp (steeply rising) or a core (flattened) (Navarro et al., 1996; Spekkens et al., 2005). N -body simulations using Λ CDM cosmology suggest that in the absence of baryonic (standard) physics, cold DM halos follow the cuspy Navarro-Frenk-White density (NFW) profile (Navarro et al., 1997), which is characterized by a steep rise in the density $\rho \propto r^{-1}$ at small 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* (Oh et al., 2011; 2015). Potential solutions to the core-cusp problem range from stellar feedback which eject baryons and flatten the DM central density profile (Navarro et al., 1996; Read & Gilmore, 2005; Mashchenko et al., 2006; Pontzen & Governato, 2012) to alternative DM models like self-interactions (Lovell et al., 2012; Elbert et al., 2015).

DM density profiles in dwarf galaxies are traditionally inferred using observations of the line-of-sight velocities and angular positions of stars gravitationally bound to these systems. In particular, Jeans equation relates the velocity dispersions of tracer stars to the gravitational potential of

¹Department of Physics, Massachusetts Institute of Technology, Cambridge, USA ²The NSF AI Institute for Artificial Intelligence and Fundamental Interactions ³Department of Physics, Harvard University, Cambridge, MA 02138, USA. Correspondence to: Tri Nguyen <tnGuy@mit.edu>.

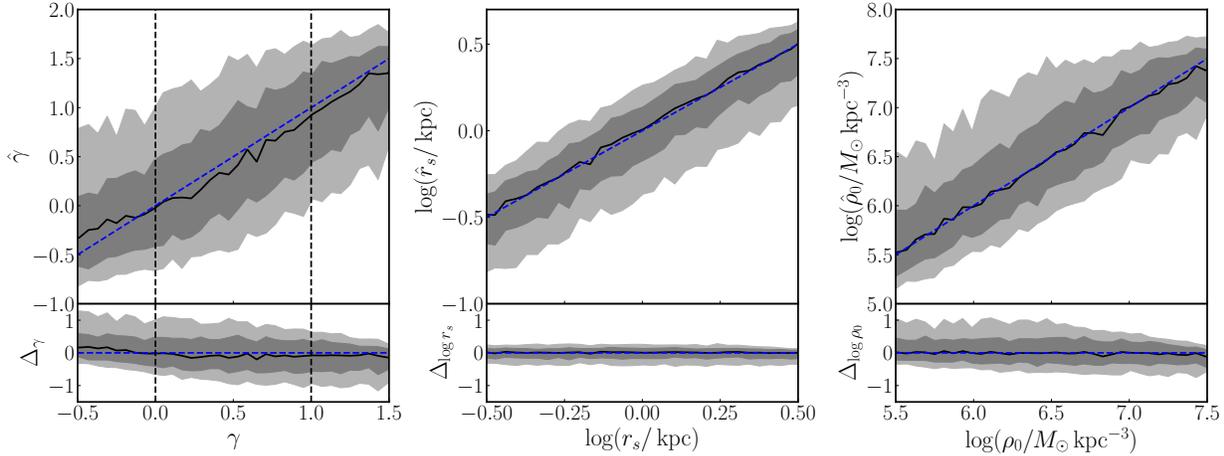


Figure 1. A comparison between the predicted and the true values of the DM parameters on 10,000 test galaxies. For each galaxy, the predicted parameters are taken to be the marginal medians of the joint posterior and then sorted into bins based on their true values. The median (solid black line), middle-68% percentile (dark gray band), and middle-95% (light gray band) containment regions of each bin are shown. The dashed blue line denotes where the predicted values are equal to the true values. The dashed vertical black lines in the left panel indicate the central slope of a cored ($\gamma = 0$) and cuspy ($\gamma = 1$) DM profile. The bottom row shows the prediction error $\Delta_x = \hat{x} - x$ for a parameter x .

the system (Jeans, 1915; Bonnavard et al., 2015).

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. El-Badry et al. (2017); Genina et al. (2020); Chang & Necib (2021)). For example, Jeans modeling assumes that the system is in dynamical equilibrium, which has recently been shown not to hold, given the active merger history of the Milky Way (see Helmi (2020) for a review). Assumptions such as isotropy of the gravitating system are also often required in order to enable a tractable analysis. Finally, by relying on a lower-dimensional description of the data through second moments of stellar velocities, inference based on Jeans modeling is likely to lose some of the salient information available in observations. In particular, leveraging the full phase-space correlation structure of stellar kinematics may further inform the latent DM density profile and lead to stronger constraints.

In this paper, we introduce a new machine learning-based approach for linking observed stellar properties to the DM profiles of dwarf galaxies. Our method is based on forward modeling of simulated dwarf galaxy systems, extracting representative features from these high-dimensional relational datasets using graph neural networks, and performing simulation-based inference for simultaneously extracting the spatial profiles associated with the DM and stellar components of the dwarf galaxy. The method aims to overcome some of the limitations associated with traditional Jeans modeling, and we demonstrate some of its advantages in terms of speed, robustness, and flexibility over established

approaches.

2. 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 graph-processing neural network and inference procedure.

2.1. Datasets and the forward model

In this proof-of-principle exposition, our forward model is fully specified by the joint distribution function (DF) 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 that we wish to infer. We use the public code `StarSampler`¹ to generate simulated realizations of stellar kinematics (6-D position and velocity phase-space components) from the forward model. `StarSampler` uses importance sampling to draw from the DF of positions and velocities of tracer stars in a given DM potential.

We model the DM profile using the generalized Navarro–Frenk–White (gNFW) profile (Navarro et al., 1997)

$$\rho_{\text{DM}}^{\text{gNFW}}(r) = \rho_0 \left(\frac{r}{r_s} \right)^{-\gamma} \left(1 + \frac{r}{r_s} \right)^{-(3-\gamma)}, \quad (1)$$

which depends on 3 free parameters: the central density ρ_0 , the scale radius r_s , and the asymptotic inner slope γ .

¹<https://github.com/maoshen1/StarSampler>

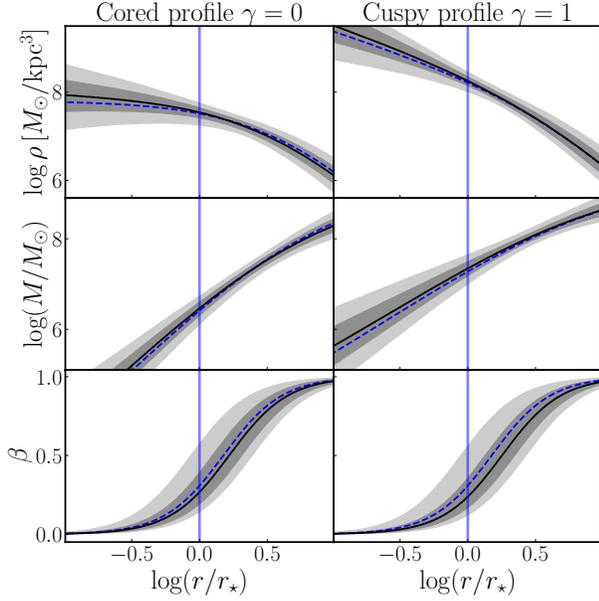


Figure 2. Example inferred posteriors of the density profile (top row), enclosed mass profile (middle row), and velocity anisotropy profile (bottom row) for dwarf galaxies with a cored DM profile (left) and a cuspy DM profile (right). The dashed blue line is the true profile, while the black line and two gray bands represent the median, 68%, and 95% containment regions.

$\gamma = 1$ corresponds to the cuspy NFW profile motivated by simulations, while $\gamma = 0$ corresponds to a pure DM core. Being able to robustly distinguish between the two possibilities would offer a path towards resolution of the core-cusp discrepancy, which is one of the high-level goals of our study.

We assume a stellar density $\nu(r)$ that follows the a 3-D Plummer profile (Plummer, 1911):

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

where L is the total luminosity and r_* is the scale length. We also introduce a velocity anisotropy profile $\beta(r)$ in order to test the performance of our method under deviations from sphericity as expected in these systems. $\beta(r)$ is described by the equation $\beta(r) = r^2/(r^2 + r_a^2)$ from Osipkov (1979); Merritt (1985) and has an additional parameter r_a which describes the transition from an isotropic velocity dispersion at small radii to a radially-biased dispersion at large radii.

In total, our model has 3 DM parameters (ρ_0, r_s, γ) and 2 stellar parameters (r_*, r_a). Assuming the gravitational influence from stars is negligible compared to the DM potential, the model is independent of the total luminosity L in Eq. 2. We sample the central density ρ_0 and scale radius r_s log-uniformly distribution from $[10^5, 10^8] M_\odot/\text{kpc}^3$ and

$[-1, 5]$ kpc and central slope γ uniformly from $[-1, 2]$. This implicitly sets the prior distributions of the predicted parameters in our Bayesian inference pipeline. Because the DM and stellar parameters are correlated, we uniformly sample r_* from $[0.2, 1] \times r_s$, and r_a from $[0.5, 2] \times r_*$.

We generate 80,000 training samples, 10,000 validation samples, and 10,000 test samples. Each sample contains the 3-D positions and 3-D velocities of tracer stars with respect to the center of a dwarf galaxy with the above parameters. The number of stars in each galaxy is sampled from a Poisson distribution with a mean of 100 stars, which roughly matches the number of stars typically observed in bright dwarf galaxies (Simon & Geha, 2007; Mateo et al., 2008; Walker et al., 2009a;b).

2.2. Data preprocessing and graph construction

We initially pre-process our datasets by adding projection effect and measurement errors in order to reflect realistic dwarf galaxy observations. For each sample, we randomly choose a line-of-sight axis and project the galaxy on the 2-D plane perpendicular to the axis. We then derive the 2-D projected coordinates (X, Y) and the line-of-sight velocities v_{los} for each star. To study the validity of the method before the inclusion of large measurement errors, we assume a Gaussian velocity measurement error of $\Delta v = 0.1$ km/s and add samples from the noise model to the intrinsic line-of-sight velocity of each star. We do not assume any positional uncertainty since the angular position measurement error for stars in dwarf galaxies is typically negligible.

A natural way to represent the stellar kinematic data is 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 $|\mathcal{V}| = N_{\text{stars}}$ individual stars, \mathcal{E} is a set of edges, and $A \in \mathbb{R}^{N_{\text{stars}} \times N_{\text{stars}}}$ is an adjacency matrix describing the weights of connections between vertices. This representation is well-suited for our purposes because the stars in a dwarf galaxy have no intrinsic ordering, and it is advantageous to explicitly encode relationships between stars (e.g., nearby stars might be expected to have similar intrinsic velocities due to being in a similar local gravitational environment).

In our analysis, each node represents a star, with the node features being its error-convolved line-of-sight velocity \tilde{v}_{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 is expected to enhance the simulation-efficiency of the downstream inference task.

To determine the graph edges \mathcal{E} , we use (X, Y) to calculate pair-wise distances between all stars. For each star, we connect it to the k -nearest stars including itself (i.e. self-loops). Since the edges are assumed to be undirected, each star can

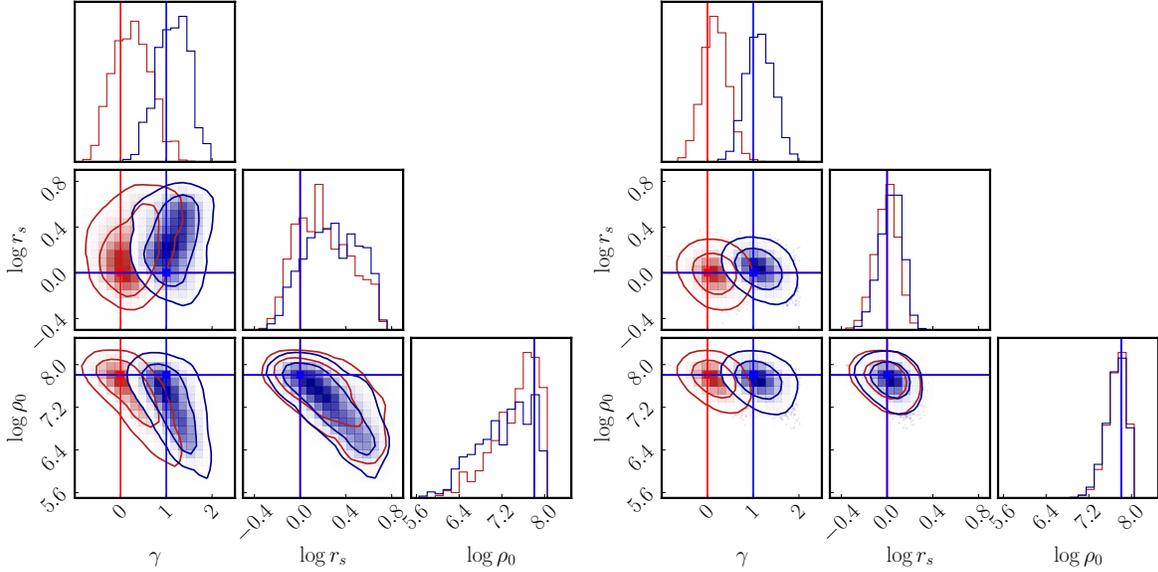


Figure 3. Example corner plots of the posterior DM parameters from Jeans dynamical modeling (left) and our method (right) on two test galaxies with cored DM profile (red) and cuspy DM profile (blue). Both galaxies have the same central slope ρ_0 and scale radius r_s . Similarly to the training dataset, each galaxy has about 100 stars and a line-of-sight velocity error of $\Delta v = 0.1$ km/s. The contour lines show the 68% and the 95% containment regions. As evidenced, our method provides a stronger constraint on the DM parameters and is able to distinguish more cleanly between a cored and cuspy profile.

be connected to more than k other stars (including itself). A higher value of k increases the number of edges, which provides more neighboring information but also increases computational and memory costs. After experimenting with different values of k , we found $k = 20$ to provide a good balance between downstream prediction accuracy and computational cost. Finally, we do not include edge weights in our graph, but note that we have experimented with a variety of weighting schemes (including attention-based learned weights (Veličković et al., 2017) as well as weights exponentially decaying with inter-star distance) and found them to perform similarly in downstream inference to the unweighted case. We will explore further variations on graph construction, including different weighted adjacency schemes and node features, in future work.

2.3. Neural network architecture and optimization

We train a graph neural network (GNN) $g_\varphi : \mathcal{G} \rightarrow \mathbb{R}^{N_{\text{feat}}}$ in order to extract N_{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 neural network. The feature-extraction network consists of 5 graph-convolutional layers based on convolutions using a basis of Chebychev polynomials in the spectral domain (Defferrard et al., 2016). 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 which projects the output

onto a set of $N_{\text{feat}} = 128$ summaries $g_\varphi(x)$.

The joint posterior $\hat{p}_\phi(\theta | g_\varphi(x))$ of the 5 parameters of interest θ characterizing the DM and stellar profiles is modeled using a normalizing flow (Papamakarios et al., 2019; Rezende & Mohamed, 2015; Papamakarios et al., 2017), which are 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 training loss.

The GNN and normalizing flow parameters $\{\varphi, \phi\}$ are optimized simultaneously on the 80,000 simulated samples using the AdamW optimizer (Kingma & Ba, 2014; Loshchilov & Hutter, 2019) with a learning rate of 5×10^{-4} . At the end of each epoch, we evaluate the loss on the 10,000 validation samples and reduce the learning rate by a factor of 10 if no improvement is seen after a few epochs. We stop the training process if the validation loss does not improve after 10 epochs. Our model takes about 1 hour to converge on an NVIDIA Tesla V100 GPU.

3. Results and discussion

We apply our method to the 10,000 test galaxies and summarize our results in Fig. 1. For each galaxy, we condition the trained flow on features extracted using the GNN and

draw 10,000 samples from the joint DM and stellar posterior. Then we compute the marginal medians as the predicted parameters and sort them into bins based on their true values. Fig. 1 shows the median (solid black line), middle-68% (dark gray bands), and middle-95% (light gray bands) containment regions for each bin of the DM parameters. In general, our method successfully recovers individual DM parameters. We note that predictions of the inner slope γ are slightly biased towards the ends of the prior distribution (which spans from $[-1, 2]$ for γ) due to edge effects. These biases are well within the modeled uncertainty and thus should not affect the overall conclusions of our study.

To test whether our method can separate a cored and ($\gamma = 0$) and a cuspy ($\gamma = 1$) DM profile, in Fig. 2 and Fig. 3, we show the inferred posteriors on two test dwarf galaxies with the same DM central density, scale radius, and stellar profile, but with different inner density slopes ($\gamma = 0$ and $\gamma = 1$). Fig. 2 shows the posteriors on the recovered density profile (top row), enclosed mass profile (middle row), and anisotropy profile (bottom row). The middle-68% and 95% containment regions are shown as gray bands, and the true latent profiles are shown with the dashed blue lines. It can be seen that the method is able to successfully reconstruct the density, mass, and anisotropy profiles at both small and large radii.

In Fig. 3, we show the corner plot of 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 95% containment regions as the contour lines. For Jeans analysis, we apply a similar procedure to Chang & Necib (2021) and use the `dynesty` module (Speagle, 2020) to sample the joint posteriors. As compared to the Jeans analysis, our method provides a stronger constraint on the DM posterior parameters and distinguishes more clearly between galaxies with cored and cuspy DM profiles.

To conclude, we found that our method based on graph neural networks and simulation-based inference outperforms established methods which use Jeans modeling in speed as well as constraining power. The latter is due to the fact that our method incorporates more information about the phase-space of bound stars. Additionally, the method simultaneously models the stellar light profile and does not require this to be done in advance as is typical for traditional approaches. While we used simulations of anisotropic spherical systems as a proof-of-concept here, in future work we plan to incorporate non-equilibrium dynamics using cosmological simulations of isolated dwarfs as well as satellites of Milky Way-like systems, which would take into account baryonic effects like tidal disruption (Garrison-Kimmel et al., 2019).

Software and Data

This research made use of the `dynesty` (Speagle, 2020), `IPython` (Perez & Granger, 2007), `Jupyter` (Kluyver et al., 2016), `Matplotlib` (Hunter, 2007), `NumPy` (Harris et al., 2020), `PyTorch` (Paszke et al., 2019), `PyTorch Geometric` (Fey & Lenssen, 2019), `PyTorch Lightning` (Falcon et al., 2020), and `SciPy` (Virtanen et al., 2020) software packages.

Acknowledgements

This work was performed in part at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611. This work is supported by the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, <http://iaifi.org/>). 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 Number DE-SC0012567.

References

- Bonnivard, V., Combet, C., Maurin, D., and Walker, M. G. Spherical Jeans analysis for dark matter indirect detection in dwarf spheroidal galaxies - impact of physical parameters and triaxiality. *MNRAS*, 446:3002–3021, January 2015. doi: 10.1093/mnras/stu2296.
- Chang, L. J. and Necib, L. Dark matter density profiles in dwarf galaxies: linking Jeans modelling systematics and observation. *Monthly Notices of the Royal Astronomical Society*, 507(4):4715–4733, 08 2021. ISSN 0035-8711. doi: 10.1093/mnras/stab2440. URL <https://doi.org/10.1093/mnras/stab2440>.
- Defferrard, M., Bresson, X., and Vandergheynst, P. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. *arXiv e-prints*, art. arXiv:1606.09375, June 2016.
- El-Badry, K., Wetzel, A. R., Geha, M., Quataert, E., Hopkins, P. F., Kereš, D., Chan, T. K., and Faucher-Giguère, C.-A. When the Jeans Do Not Fit: How Stellar Feedback Drives Stellar Kinematics and Complicates Dynamical Modeling in Low-mass Galaxies. *ApJ*, 835(2):193, Feb 2017. doi: 10.3847/1538-4357/835/2/193.
- Elbert, O. D., Bullock, J. S., Garrison-Kimmel, S., Rocha, M., Oñorbe, J., and Peter, A. H. G. Core formation in dwarf haloes with self-interacting dark matter: no fine-tuning necessary. *MNRAS*, 453(1):29–37, 08 2015. ISSN 0035-8711. doi: 10.1093/mnras/stv1470. URL <https://doi.org/10.1093/mnras/stv1470>.

- Falcon, W. et al. Pytorchlightning/pytorch-lightning: 0.7.6 release, May 2020. URL <https://doi.org/10.5281/zenodo.3828935>.
- Fey, M. and Lenssen, J. E. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.
- Garrison-Kimmel, S., Hopkins, P. F., Wetzel, A., Bullock, J. S., Boylan-Kolchin, M., Kereš, D., Faucher-Giguère, C.-A., El-Badry, K., Lamberts, A., Quataert, E., and Sanderson, R. The Local Group on FIRE: dwarf galaxy populations across a suite of hydrodynamic simulations. , 487(1):1380–1399, July 2019. doi: 10.1093/mnras/stz1317.
- Genina, A., Read, J. I., Frenk, C. S., Cole, S., Benítez-Llambay, A., Ludlow, A. D., Navarro, J. F., Oman, K. A., and Robertson, A. To β or not to β : can higher order Jeans analysis break the mass–anisotropy degeneracy in simulated dwarfs? *Monthly Notices of the Royal Astronomical Society*, 498(1):144–163, 08 2020. ISSN 0035-8711. doi: 10.1093/mnras/staa2352. URL <https://doi.org/10.1093/mnras/staa2352>.
- Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., et al. Array programming with numpy. *Nature*, 585(7825):357–362, 2020.
- Helmi, A. Streams, Substructures, and the Early History of the Milky Way. , 58:205–256, August 2020. doi: 10.1146/annurev-astro-032620-021917.
- Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.
- Jeans, J. H. On the theory of star-streaming and the structure of the universe. *MNRAS*, 76:70–84, December 1915. doi: 10.1093/mnras/76.2.70.
- Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- Kluyver, T. et al. Jupyter notebooks - a publishing format for reproducible computational workflows. In *ELPUB*, 2016.
- Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL <https://openreview.net/forum?id=Bkg6RiCqY7>.
- Lovell, M. R., Eke, V., Frenk, C. S., Gao, L., Jenkins, A., Theuns, T., Wang, J., White, S. D. M., Boyarsky, A., and Ruchayskiy, O. The haloes of bright satellite galaxies in a warm dark matter universe. *MNRAS*, 420(3):2318–2324, 02 2012. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2011.20200.x. URL <https://doi.org/10.1111/j.1365-2966.2011.20200.x>.
- Mashchenko, S., Couchman, H. M. P., and Wadsley, J. The removal of cusps from galaxy centres by stellar feedback in the early Universe. *Nature*, 442(7102):539–542, August 2006. doi: 10.1038/nature04944.
- Mateo, M., Olszewski, E. W., and Walker, M. G. The Velocity Dispersion Profile of the Remote Dwarf Spheroidal Galaxy Leo I: A Tidal Hit and Run? *ApJ*, 675(1):201–233, March 2008. doi: 10.1086/522326.
- Merritt, D. Spherical stellar systems with spheroidal velocity distributions. *AJ*, 90:1027–1037, June 1985. doi: 10.1086/113810.
- Navarro, J. F., Eke, V. R., and Frenk, C. S. The cores of dwarf galaxy haloes. *MNRAS*, 283(3):L72–L78, 12 1996. ISSN 0035-8711. doi: 10.1093/mnras/283.3.L72. URL <https://doi.org/10.1093/mnras/283.3.L72>.
- Navarro, J. F., Eke, V. R., and Frenk, C. S. The cores of dwarf galaxy haloes. *MNRAS*, 283(3):L72–L78, December 1996. doi: 10.1093/mnras/283.3.L72.
- Navarro, J. F., Frenk, C. S., and White, S. D. M. A universal density profile from hierarchical clustering. *ApJ*, 490(2):493–508, dec 1997. doi: 10.1086/304888. URL <https://doi.org/10.1086/304888>.
- Oh, S.-H., Brook, C., Governato, F., Brinks, E., Mayer, L., de Blok, W. J. G., Brooks, A., and Walter, F. The Central Slope of Dark Matter Cores in Dwarf Galaxies: Simulations versus THINGS. *AJ*, 142(1):24, July 2011. doi: 10.1088/0004-6256/142/1/24.
- Oh, S.-H., Hunter, D. A., Brinks, E., Elmegreen, B. G., Schruha, A., Walter, F., Rupen, M. P., Young, L. M., Simpson, C. E., Johnson, M. C., Herrmann, K. A., Ficut-Vicas, D., Cigan, P., Heesen, V., Ashley, T., and Zhang, H.-X. High-resolution Mass Models of Dwarf Galaxies from LITTLE THINGS. *AJ*, 149(6):180, June 2015. doi: 10.1088/0004-6256/149/6/180.
- Osipkov, L. P. Spherical systems of gravitating bodies with an ellipsoidal velocity distribution. *Pisma v Astronomicheskii Zhurnal*, 5:77–80, February 1979.
- Papamakarios, G., Pavlakou, T., and Murray, I. Masked autoregressive flow for density estimation. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, pp. 2335–2344, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964. URL <https://papers.nips.cc/paper/2017/hash/6c1da886822c67822bcf3679d04369fa-Abstract.html>.

- Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., and Lakshminarayanan, B. Normalizing flows for probabilistic modeling and inference. 2019.
- Paszke, A. et al. Pytorch: An imperative style, high-performance deep learning library. In Wallach, H., Laroche, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019. URL <http://papers.nips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
- Perez, F. and Granger, B. E. IPython: A System for Interactive Scientific Computing. *Computing in Science and Engineering*, 9(3):21–29, Jan 2007. doi: 10.1109/MCSE.2007.53.
- Plummer, H. C. On the problem of distribution in globular star clusters. *MNRAS*, 71:460–470, March 1911. doi: 10.1093/mnras/71.5.460.
- Pontzen, A. and Governato, F. How supernova feedback turns dark matter cusps into cores. *MNRAS*, 421(4):3464–3471, April 2012. doi: 10.1111/j.1365-2966.2012.20571.x.
- Read, J. I. and Gilmore, G. Mass loss from dwarf spheroidal galaxies: the origins of shallow dark matter cores and exponential surface brightness profiles. *MNRAS*, 356(1): 107–124, January 2005. doi: 10.1111/j.1365-2966.2004.08424.x.
- Rezende, D. J. and Mohamed, S. Variational inference with normalizing flows. In Bach, F. R. and Blei, D. M. (eds.), *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pp. 1530–1538, 2015. URL <http://proceedings.mlr.press/v37/rezende15.html>.
- Simon, J. D. and Geha, M. The Kinematics of the Ultrafaint Milky Way Satellites: Solving the Missing Satellite Problem. *ApJ*, 670(1):313–331, November 2007. doi: 10.1086/521816.
- Speagle, J. S. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. , 493(3):3132–3158, April 2020. doi: 10.1093/mnras/staa278.
- Spekkens, K., Giovanelli, R., and Haynes, M. P. The Cusp/Core Problem in Galactic Halos: Long-Slit Spectra for a Large Dwarf Galaxy Sample. *AJ*, 129(5):2119–2137, May 2005. doi: 10.1086/429592.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. Graph Attention Networks. *arXiv e-prints*, art. arXiv:1710.10903, October 2017.
- Virtanen, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 2020. doi: <https://doi.org/10.1038/s41592-019-0686-2>.
- Walker, M. G., Mateo, M., and Olszewski, E. W. Stellar Velocities in the Carina, Fornax, Sculptor, and Sextans dSph Galaxies: Data From the Magellan/MMFS Survey. *ApJ*, 137:3100–3108, February 2009a. doi: 10.1088/0004-6256/137/2/3100.
- Walker, M. G., Mateo, M., Olszewski, E. W., Peñarrubia, J., Evans, N. W., and Gilmore, G. A Universal Mass Profile for Dwarf Spheroidal Galaxies? *ApJ*, 704(2):1274–1287, October 2009b. doi: 10.1088/0004-637X/704/2/1274.
- White, S. D. M. and Rees, M. J. Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. , 183:341–358, May 1978. doi: 10.1093/mnras/183.3.341.