Weighing the Milky Way and Andromeda galaxies with artificial intelligence

Pablo Villanueva-Domingo,[1,*] Francisco Villaescusa-Navarro,[2,3,†] Shy Genel,[2,4] Daniel Anglés-Alcázar,[5,2] Lars Hernquist,[6] Federico Marinacci,[7] David N. Spergel,[2,3] Mark Vogelsberger,[8] and Desika Narayanan[9,10]

[1]*Instituto de Física Corpuscular (IFIC), CSIC-Universitat de València, E-46980 Paterna, Spain*
[2]*Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, New York 10010, USA*
[3]*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, New Jersey 08544, USA*
[4]*Columbia Astrophysics Laboratory, Columbia University, New York, New York 10027, USA*
[5]*Department of Physics, University of Connecticut, 196 Auditorium Road, U-3046, Storrs, Connecticut 06269-3046, USA*
[6]*Center for Astrophysics—Harvard and Smithsonian, 60 Garden St, Cambridge, Massachusetts 02138, USA*
[7]*Dipartimento di Fisica e Astronomia "Augusto Righi," Università di Bologna, via Gobetti 93/2, 40129 Bologna, Italy*
[8]*Kavli Institute for Astrophysics and Space Research, Department of Physics, MIT, Cambridge, Massachusetts 02139, USA*
[9]*Department of Astronomy, University of Florida, Gainesville 32611, Florida, USA*
[10]*University of Florida Informatics Institute, 432 Newell Drive, CISE Bldg E251, Gainesville 32603, Florida, USA*

(Received 9 December 2021; accepted 8 March 2023; published 2 May 2023)

We present new constraints on the masses of the halos hosting the Milky Way and Andromeda galaxies derived using graph neural networks. Our models, trained on 2,000 state-of-the-art hydrodynamic simulations of the CAMELS project, only make use of the positions, velocities and stellar masses of the galaxies belonging to the halos, and are able to perform likelihood-free inference on halo masses while accounting for both cosmological and astrophysical uncertainties. Our constraints are in agreement with estimates from other traditional methods, within our derived posterior standard deviation.

DOI: 10.1103/PhysRevD.107.103003

I. INTRODUCTION

The ΛCDM model describes how nonlinear gravitational evolution amplifies the small dark matter perturbations after inflation creating larger and deeper gravitational potential wells where gas can collapse, cool down, and form stars and galaxies [1]. The observed rotation curves of galaxies indicate the existence of an additional mass beyond the one that can be accounted for from luminous matter, suggesting that galaxies are embedded in larger objects called halos [2–4]. Determining the mass of the halo hosting our galaxy or that of our neighbors would provide us with a better understanding of the dark matter-baryon relation [5]. Furthermore, knowing the total mass of our galaxy and its companions will help us to determine their fate [6,7].

In [8] we developed a new machine learning method that uses graph neural networks (GNNs) to perform likelihood-free inference of halo masses from the positions, velocities, and internal properties of the galaxies they host. Our model does not make any assumption about the state of the system, and since it was trained on galaxies from simulations of the Cosmology and Astrophysics with MachinE Learning Simulations (CAMELS) project [9], automatically marginalizes over the cosmological and astrophysical uncertainties as modeled in CAMELS. We also showed that the model predictions are robust, i.e., the outcome weakly depends on the particular type of simulation used to train the GNN. In this work we use those networks to infer the individual masses of the halos hosting the Milky Way (MW) and Andromeda (M31) galaxies, being, to our knowledge, the first time that artificial intelligence is applied to this task.[1]

*pablo.villanueva.domingo@gmail.com
†fvillaescusa@flatironinstitute.org

[1]Note, however, that the joint mass of the MW + M31 system has been previously inferred using machine learning methods [10].

 © 2023 American Physical Society

II. MODEL

Our model works as follows. Given a halo, we first select all the galaxies belonging to it. Next, we create a mathematical graph, where each of the halo galaxies (including both central and satellites) are considered as nodes that are connected by edges depending on their relative distances. The nodes can have several properties associated with them, such as positions, velocities, and stellar masses of the galaxies they represent.

GNNs are deep learning architectures appropriate to deal with irregular data, like the graphs constructed above [11–13]. They exhibit many advantages over traditional convolutional neural networks (CNNs), and given the sparsity, structure and irregularity of the data we are working with, GNNs are more suited for this problem. They employ the so-called message passing scheme, where the node features of neighbors are aggregated to infer global quantities of the graph. In that way, GNNs exploit neighborhood relations and locality. These models are by construction permutation invariant, and we also perform data augmentation to enforce the network to be rotationally invariant. The GNNs are trained to perform likelihood-free inference of the halo mass, in order to output the posterior mean and standard deviation without making any assumption on the form of the posterior distribution [14]. We use M_{200c}, the mass within a virial radius with a density 200 times the critical density, as the definition of halo mass. We refer the reader to [8] for further details on the architecture and training procedure.

We train the networks on halos/galaxies from the CAMELS simulations [9,15,16], containing two different simulation suites that were run with two distinct codes employing different subgrid physics models: (i) AREPO[2] [17] and the IllustrisTNG model [18–20], and (ii) GIZMO[3] [21] and the SIMBA model [22]. These models differ in their treatment of astrophysical processes such as supernova and active galactic nuclei (AGN) feedback or black hole growth. Each suite (IllustrisTNG and SIMBA) contains 1000 simulations, each of them having different values of two cosmological parameters ($\Omega_{\rm m}$ and σ_8) and four astrophysical parameters controlling the efficiency of feedback from supernovae and AGNs.[4] The halos and subhalos have been identified with the FOF + SUBFIND algorithm [23]. In the simulations, galaxies are defined as subhalos that contain more than 10 star particles. In this paper we only consider galaxy properties that can be accessed via observations[5]: (i) positions, (ii) modulus of the velocity v, and (iii) stellar mass of each galaxy M_*. Positions and velocities of galaxies are defined relative to the rest frame of the central galaxy, about its origin. In this work we have trained four different networks: (1) using IllustrisTNG galaxies including velocities, (2) using IllustrisTNG galaxies without velocities, (3) using SIMBA galaxies including velocities, and (4) using SIMBA galaxies without velocities. We note that since the networks are trained employing galaxies having different cosmological and astrophysical models, the network learns to marginalize over these. We emphasize that in this setup we are considering all CAMELS models equally likely. In other words, we are assuming a flat prior on the value of the cosmological and astrophysical parameters.

III. DATA

With the networks trained on simulated galaxies we apply them to real data. The inputs to the model are the positions, velocities, and stellar masses of the galaxies belonging to the halos hosting the MW and M31 galaxies. The list of known satellite galaxies of the MW has grown during the last years up to several dozens thanks to observations by DES [24,25] and SDSS [26]. The PAndAS survey has also extended the number of M31 satellites above 30 [27].[6] When building our graphs we only consider galaxies with stellar masses above $\sim 10^8\ h^{-1} M_\odot$, since those are the galaxies present in the CAMELS simulations and therefore the ones used to train the GNNs.

The MW graph contains thus the Large and Small Magellanic Clouds (LMC and SMC), together with the Milky Way (MW) itself. The remaining dwarf satellites are much less massive [30]. Stellar masses are extracted from [31,32], while velocities for the Magellanic clouds are computed from radial velocities and proper motions from [33–37]. Positions in the galactic coordinate frame have been transformed to Cartesian galactocentric frame making use of astropy[7] [38,39].

The Andromeda graph contains M31, M33 (also known as the Triangulum galaxy), M32 (also known as Le Gentil), and M110 (also known as NGC 205). Note that the membership of M33 to the M31 halo is still under discussion, but the general consensus is that it is a satellite [40–42]. We take the positions and stellar masses of Andromeda and its satellites from [28]. The galactocentric velocities for M31 and M33 are extracted from [43], using $\vec{v}_{\rm 3D,M31} = (34, -123, -19)$ km/s from their DR2 + HST weighted average. Note, however, that there is not complete agreement about the tangential velocity of M31, which strongly relies on the specific approach followed [40,43,44]. 3D velocities for M32 and M110 have not been derived in the literature, since their proper motions are difficult to measure, given their proximity

[2] https://arepo-code.org/.

[3] http://www.tapir.caltech.edu/ phopkins/Site/GIZMO.html.

[4] This corresponds to the CAMELS LH set [8,9].

[5] We note that the models in [8] also include the stellar half-mass radius as a galactic feature. However, since it may be difficult to estimate it observationally for some satellites, we do not include it in our analysis. Excluding this feature does not dramatically decrease the accuracy, presenting mean relative errors (in the logarithm of the mass) between ∼1%–2%.

[6] See Refs. [28,29] for comprehensive compilations of the MW and M31 satellite galaxy data.

[7] https://www.astropy.org.

TABLE I. We use graph neural networks (GNNs) to infer the masses of the halos hosting the Milky Way (MW) and the Andromeda (M31) galaxies. This table shows the galaxies taken into account and their properties for both cases. The input data comprise the position in Cartesian galactocentric coordinates, the modulus of the velocity in the central galaxy rest frame (either MW or M31), and the stellar mass of each galaxy. Velocities marked with an asterisk are estimated by multiplying their radial components by $\sqrt{3}$.

Halo	Galaxy	Position (kpc)	Velocity (km/s)	Stellar mass ($M_\odot$)	References
MW	MW	(0,0,0)	0.0	5×10^{10}	[31]
	LMC	$(-0.6, -41.3, -27.1)$	322.1	2.7×10^9	[32–35]
	SMC	$(15.8, -37.3, -43.3)$	237.4	3.1×10^8	[32,35–37]
M31	M31	$(-386.0, 622.7, -287.3)$	0.0	1.0×10^{11}	[28,43,45,46]
	M33	$(-485.9, 500.6, -419.0)$	257.6	2.9×10^9	[28,43]
	M110	$(-401.4, 661.0, -295.6)$	116.1*	3.3×10^8	[28]
	M32	$(-395.5, 638.4, -300.6)$	38.1*	3.2×10^8	[28]

to M31; only radial velocities have been estimated for these galaxies. Given that minor variations in the velocity of these small satellites are not expected to have a significant impact on the results, since only the modulus of the relative velocity is required rather than specific 3D values, and as long as other satellites also present large uncertainties in their velocities, we set their velocities to their radial component in the M31 frame multiplied by a factor of $\sqrt{3}$. We have verified that the precise value of the velocities of M32 and M110 does not have a major impact on the determination of the M31 mass. Table I presents a summary of the features and sources considered for each halo. Note that we treat the MW and M31 datasets separately, i.e., inferring masses for each halo individually.

IV. RESULTS

Several distinct methods have been used to infer the mass of the MW halo: (i) from Local Group dynamics [47–51], (ii) modelling stellar tidal streams [52,53], (iii) rotation curve data [54–58], (iv) the escape speed of the galaxy [59–61], (v) properties and dynamics of satellite galaxies [41,62–69], and (vi) halo globular cluster kinematics [70,71]. See, e.g., [56,72], for recent comprehensive compilations of estimates of the MW mass using different approaches. Some of these methods have also been devoted to estimate the mass of the Andromeda halo, such as (i) from the Local Group dynamics [48–51], (ii) stellar streams [73], (iii) rotation curve data [54], (iv) the escape velocity of planetary nebulae [74], (v) satellite phenomenology [41,75–78], (vi) globular clusters [79], and (vii) the distribution of its stellar mass [80]. Other works have inferred instead the sum of the MW and M31 total masses [10,40,81–83]. The colored points in Fig. 1 show the constraints derived for the MW (left) and M31 (right) halos.

The black points of Fig. 1 show instead the bounds derived for the MW and M31 halo masses from the four GNNs mentioned using the two training simulation suites and whether they incorporate peculiar velocities as features. Table II summarizes these results. As can be seen, the predictions from models employing only positions and stellar mass (shown as black ♠ points) are generally in good agreement with most of the previous literature, within the confidence limits. The estimated uncertainties are slightly larger than those from other traditional techniques, mainly due to the few features considered and given that the GNNs have been trained on simulations varying cosmological and astrophysical parameters. Predictions of the M31 halo mass taking into account also velocities (shown as black ♣ points in the right panel of Fig. 1) are also consistent with standard estimations, but show smaller uncertainties, since the model is more precise when employing an additional feature. It is worth it to highlight that results from models trained with different simulation suites, IllustrisTNG and SIMBA, are compatible in both the central values and the confidence region, illustrating the robustness of the method.

However, we find that the predictions for the MW employing a GNN where galaxy velocities are included (shown as black ♣ points in the left panel of Fig. 1) lead to a larger halo mass. While these networks are more accurate, these predictions are in worse agreement with most of the literature estimates (although still consistent with some estimates from satellite and Local Group dynamics). We believe this result can be attributed to the fact that the velocity of the LMC is relatively large. To see this, we show in Fig. 2 the velocities versus stellar masses of galaxies within MW-like halos (with total masses between 0.8×10^{12} and $1.5 \times 10^{12} M_\odot$) from both CAMELS suites: IllustrisTNG and SIMBA. Local Group satellite galaxies considered are also shown. It can be noted that the LMC presents a relatively large velocity compared to most CAMELS satellites of the same stellar mass. We believe that this large value of the most massive satellite would lead the GNN to predict a larger value of the halo mass. This fact is also consistent with other works in the literature employing velocities to infer the MW mass. For instance, the LMC velocity suggests a MW mass estimate larger than usual [41,84]. It has been argued that this large velocity could be due to the fact that the LMC is on its first passage about

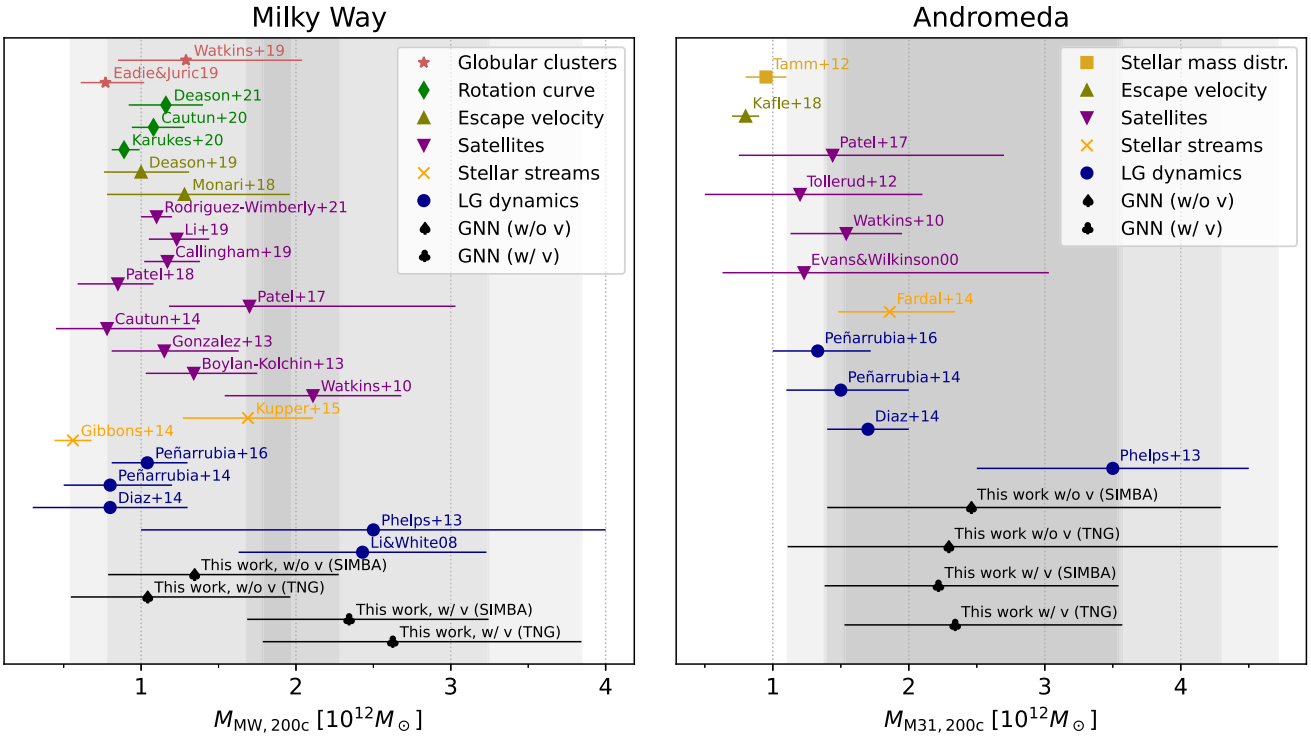

FIG. 1. The colored points with error bars show constraints on the mass (M_{200c}) of the halo hosting the Milky Way (left) and the Andromeda (right) galaxies using different methods. The black points with error bars display instead the halo masses derived from graph neural network models that are trained from CAMELS simulations with and without velocities of the galaxies. Shaded regions correspond to the uncertainty of the GNN predictions.

the MW [85]. Moreover, the velocities of LMC and SMC-like satellites are found to be larger when the Local Group environment is taken into account [63]. This goes along with the fact pointed out in previous works that the MW may behave as a somewhat unusual galaxy for some aspects, e.g., regarding its satellite population [86] or its luminosity and rotational velocity relation [87,88].

It is worth examining the robustness of the GNN predictions. Given that the networks have been trained applying random rotations to the galaxies in the halos, it is expected that the inference would be rotationally invariant. By performing 50 arbitrary rotations of the MW and M31 graphs, we have verified that our results are stable to $\lesssim$0.05 times the mean values. This shows that, effectively, the GNN outputs are robust under these transformations.

TABLE II. Constraints on the masses of the MW and M31 halos from models trained in IllustrisTNG and SIMBA simulations, either including or excluding velocities to train the models.

	Mass $(10^{12} M_\odot)$			
	IllustrisTNG		SIMBA	
Halo	w/o v	w/ v	w/o v	w/ v
MW	$1.0^{+0.9}_{-0.5}$	$2.6^{+1.2}_{-0.8}$	$1.3^{+0.9}_{-0.6}$	$2.3^{+0.9}_{-0.7}$
M31	$2.3^{+2.4}_{-1.2}$	$2.3^{+1.2}_{-0.8}$	$2.5^{+1.8}_{-1.0}$	$2.2^{+1.3}_{-0.8}$

On the other hand, it is pertinent to check how sensitive the outputs are to the number of satellites considered. The MW and M31 predictions can increase by up to ∼15% when the smallest satellites, the SMC and M32, respectively, are removed. If M110 is also removed from the M31 group,

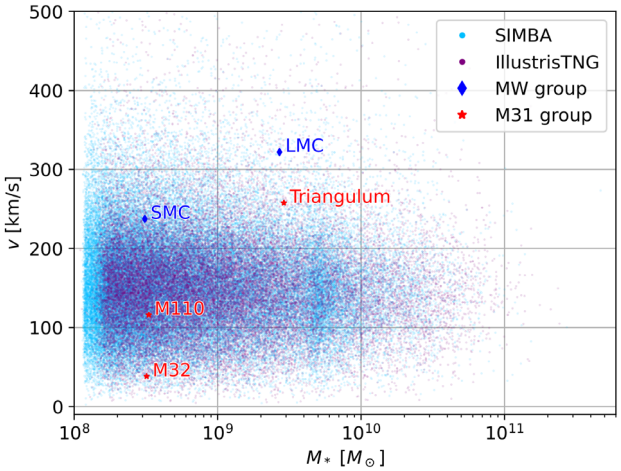

FIG. 2. Velocity v versus stellar mass M_* for satellite galaxies within MW-like halos, with total mass between 0.8×10^{12} and $1.5 \times 10^{12} M_\odot$, in the CAMELS simulations. Scatter points are extracted from the IllustrisTNG and SIMBA simulations suites. Satellite galaxies from the MW and M31 systems are also shown.

there can be an additional ∼15% increase if the velocities are employed, or a ∼5% decrease when only positions and stellar masses are considered. This change of tendency if velocities are employed or not reveals that the GNN may rely more on some satellites than others depending on the features considered, as shown in [8]. We cannot push this further, as the network is trained on halos with at least one satellite, so it would not be reliable to apply it to only the central galaxy.

Analogously, we have also checked the sensitivity of the results when the next satellites in stellar mass are also taken into account. The following satellites in order of decreasing stellar mass are the Sagittarius Dwarf Spheroidal Galaxy (Sgr dSph) and NGC 147 for the MW and M31, respectively, having both stellar masses just below the CAMELS galaxy mass resolution, around $\sim 10^8 M_\odot$ [89,90]. Considering Sgr dSph for the MW halo can lead to a ∼15% increase, while NGC 147 enhances the M31 mass up to a ∼30%. Note, however, that these predictions are not reliable and are only useful for testing purposes, since the GNNs have been trained only with galaxies above the CAMELS mass resolution. Sgr dSph and NGC 147 stellar mass values are not sampled in the training dataset and thus there is no guarantee that the GNN is able to extrapolate there. Anyway, all these tests show that the presence of satellite galaxies is highly relevant and has a non-negligible impact on the final results.

Finally, so far we have neglected the effect of the uncertainties of observational data in the error of the mass constraint. To properly evaluate their impact, we would need to compute the covariance between the different data. Here we limit to provide a simple estimate on the effect of observable uncertainties. We focus on the models trained with only positions and stellar masses (i.e., without considering velocities). Furthermore, we consider the error of the stellar mass of the central galaxy as the only source of uncertainty, as it is the most prominent expected source of error and the most relevant feature within this configuration [8]. The values considered are $M_{*,\mathrm{MW}} = (5 \pm 1) \times 10^{10} M_\odot$ for the MW [31] and $M_{*,31} = (10.3^{+2.3}_{-1.7}) \times 10^{10}\ M_\odot$ for M31 [45]. We have randomly sampled 100 values within the confidence range, and used these stellar masses to query the GNN and compute the standard deviation. Using the model trained with SIMBA simulations, the results are $M_{\mathrm{MW},200c} = (1.3 \pm 0.1) \times 10^{12} M_\odot$ for the MW and $M_{M31,200c} = (2.5 \pm 0.3) \times 10^{12}\ M_\odot$ for M31, obtaining similar errors with the model trained in IllustrisTNG. Note that these reported errors only include the observational uncertainty from the stellar mass of the central galaxy, and not those from the likelihood-free inference of the GNN (shown in Table II), which are always notably larger. These results illustrate that the main source of uncertainty comes from the likelihood-free inference approach to train the GNN rather than from observational sources. Note, however, that future applications of this method would require a proper evaluation of observable uncertainties, in order to guarantee reliable results.

V. DISCUSSION

We have applied a novel machine learning model trained on hydrodynamic simulations to predict the masses of the halos hosting the MW and M31 galaxies, obtaining results in good agreement with standard observational methods from rotation curves, Local Group dynamics, kinematic tracers, and others techniques. To our knowledge, these are the first constraints on the individual MW and M31 halo masses from artificial intelligence. The predicted masses do not make any assumptions about the underlying cosmology and astrophysical scenario, since the models have been trained to marginalize over a broad parameter space. This achievement illustrates the suitability of these models to be applied in other galactic systems. Furthermore, models trained with different simulation suites provide highly compatible results, indicating the robustness of the method. Our GNN-based method relies on few observational galactic properties, including the 3D positions, velocities, and stellar masses. When velocities are considered as features, predictions for the MW are slightly larger than those using only positions and stellar mass. This behavior is, however, expected, given the large velocity of LMC with respect to their counterparts in simulations.

Our method presents some caveats which are worth noting. First, we only provided a simple estimate of the impact of observational errors in the galaxy properties belonging to the halos containing MW and M31. Properly accounting for these statistical uncertainties via the covariance matrix would increase the expected confidence level of our predictions. Furthermore, it is not clear which observational features would be more relevant to infer halo masses. Interpretability is one of the main limitations of deep learning. While one may think that the GNN learns to approximate a virial-like mass-velocity relation to extract the total halo mass, it is not possible for us to make strong claims on whether the network is learning such mapping. In contrast, in our previous paper [8] we made use of saliency graphs to try to visualize the important elements of the model, finding that the GNNs tend to exploit the central galaxy information, and occasionally also focus on some small satellites, but the interpretability is still unclear. Also, it is possible that employing other properties (e.g., neutral hydrogen mass, metallicity, morphology, luminosity, or different velocity components), our results may become more precise, robust, and reliable. Furthermore, we have implicitly assumed that the velocity is isotropic for some of the M31 satellites which lack velocity data, which may induce some biases in the M31 prediction. Using 3D velocities when possible would allow us to find more restrictive results. Note that we have implicitly assumed that the CAMELS training set is representative of the MW and M31 galaxies, which could not be the case, since the MW may present atypical features in some regards. Note also that the CAMELS simulations adopt a wide flat prior over the range of initial conditions and parameters

employed, which could give too much weight to some unlikely and extreme cosmological and astrophysical scenarios. Training the GNNs with more realistic and tighter priors on some parameters may also enhance the accuracy and reduce the uncertainties, as shown in [8] within a fixed cosmology, as well as limiting to models that match some observable quantity such as the stellar mass function or star-formation rate history. Moreover, increasing resolution and resolving smaller galaxies in the training simulations would allow us to include further lighter satellites, which may enhance the constraints. It is also worth it to explore variations of the GNN architecture, such as those applied to cosmological inference, making use of edge attributes [91,92] or hierarchical aggregation [93]. Moreover, the effect of nearby halos as interlopers has not been explicitly taken into account, which may impact the predictions, given that the MW and M31 are close and gravitationally bound. One could restrict the set of simulations used for training the models to those with properties similar to the MW and M31 system, either in terms of their current configuration as a pair of binary galaxies approaching each other, or alternatively having a history that matches their expected evolution. Besides, training on simulations with higher resolution will enable the inclusion of smaller and fainter satellites that can provide additional information and therefore tighten the constraints. This method is hence a proof-of-concept rather than a definitive technique, since there is still room for improvement, which is left for future work.

The implementation of the GNNs underlying this article, `HaloGraphNet`, is available on GitHub at [94,95], making use of PyTorch Geometric [96]. Details on the CAMELS simulations can be found at [97].

ACKNOWLEDGMENTS

We thank Adrian Price-Whelan, Alan W. McConnachie, and Sangmo Tony Sohn for their help with Milky Way and Andromeda satellites data. We also thank Nicolas Garavito and Vasily Belokurov for enlightening discussions. The work of P. V. D. is supported by CIDEGENT/2018/019, CPI-21-108. F. V. N. was supported by funding from the WFIRST program through NNG26PJ30C and NNN12AA01C. D. A. A. was supported in part by NSF Grants AST-2009687 and AST-2108944, and by the Flatiron Institute, which is supported by the Simons Foundation. The training of the GNNs has been carried out using GPUs from the Tiger cluster at Princeton University.

[1] R. S. Somerville and R. Davé, Annu. Rev. Astron. Astrophys. **53**, 51 (2015).
[2] V. C. Rubin, W. K. Ford, Jr., and N. Thonnard, Astrophys. J. Lett. **225**, L107 (1978).
[3] A. Bosma, The distribution and kinematics of neutral hydrogen in spiral galaxies of various morphological types, Ph.D. thesis, Groningen University, 1978.
[4] Y. Sofue and V. Rubin, Annu. Rev. Astron. Astrophys. **39**, 137 (2001).
[5] R. H. Wechsler and J. L. Tinker, Annu. Rev. Astron. Astrophys. **56**, 435 (2018).
[6] M. Cautun, A. J. Deason, C. S. Frenk, and S. McAlpine, Mon. Not. R. Astron. Soc. **483**, 2185 (2019).
[7] R. P. van der Marel, G. Besla, T. J. Cox, S. T. Sohn, and J. Anderson, Astrophys. J. **753**, 9 (2012).
[8] P. Villanueva-Domingo, F. Villaescusa-Navarro, D. Anglés-Alcázar, S. Genel, F. Marinacci, D. N. Spergel, L. Hernquist, M. Vogelsberger, R. Dave, and D. Narayanan, Astrophys. J. **935**, 30 (2022).
[9] F. Villaescusa-Navarro, D. Anglés-Alcázar, S. Genel, D. N. Spergel, R. S. Somerville, R. Dave, A. Pillepich, L. Hernquist, D. Nelson, P. Torrey *et al.*, Astrophys. J. **915**, 71 (2021).
[10] P. Lemos, N. Jeffrey, L. Whiteway, O. Lahav, N. Libeskind, and Y. Hoffman, Phys. Rev. D **103**, 023009 (2021).
[11] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, arXiv:2104.13478.
[12] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner *et al.*, arXiv:1806.01261.
[13] W. L. Hamilton, *Graph Representation Learning*, Synthesis Lectures on Artificial Intelligence and Machine Learning (Springer Cham, 2020), Vol. 14, pp. 1–159.
[14] N. Jeffrey and B. D. Wandelt, arXiv:2011.05991.
[15] F. Villaescusa-Navarro, S. Genel, D. Angles-Alcazar, L. Thiele, R. Dave, D. Narayanan, A. Nicola, Y. Li, P. Villanueva-Domingo, B. Wandelt *et al.*, Astrophys. J. Suppl. Ser. **259**, 61 (2022).
[16] F. Villaescusa-Navarro, S. Genel, D. Anglés-Alcázar, L. A. Perez, P. Villanueva-Domingo, D. Wadekar, H. Shao, F. G. Mohammad, S. Hassan, E. Moser *et al.*, arXiv:2201.01300.
[17] R. Weinberger, V. Springel, and R. Pakmor, Astrophys. J. Suppl. Ser. **248**, 32 (2020).
[18] R. Weinberger, V. Springel, L. Hernquist, A. Pillepich, F. Marinacci, R. Pakmor, D. Nelson, S. Genel, M. Vogelsberger, J. Naiman *et al.*, Mon. Not. R. Astron. Soc. **465**, 3291 (2017).
[19] A. Pillepich, V. Springel, D. Nelson, S. Genel, J. Naiman, R. Pakmor, L. Hernquist, P. Torrey, M. Vogelsberger, R. Weinberger *et al.*, Mon. Not. R. Astron. Soc. **473**, 4077 (2018).
[20] D. Nelson, V. Springel, A. Pillepich, V. Rodriguez-Gomez, P. Torrey, S. Genel, M. Vogelsberger, R. Pakmor, F. Marinacci, R. Weinberger *et al.*, Comput. Astrophys. Cosmol. **6**, 2 (2019).

[21] P. F. Hopkins, Mon. Not. R. Astron. Soc. **450**, 53 (2015).
[22] R. Davé, D. Anglés-Alcázar, D. Narayanan, Q. Li, M. H. Rafieferantsoa, and S. Appleby, Mon. Not. R. Astron. Soc. **486**, 2827 (2019).
[23] V. Springel, S. D. M. White, G. Tormen, and G. Kauffmann, Mon. Not. R. Astron. Soc. **328**, 726 (2001).
[24] A. Drlica-Wagner *et al.* (DES Collaboration), Astrophys. J. **813**, 109 (2015).
[25] T. M. C. Abbott, F. B. Abdalla, S. Allam, A. Amara, J. Annis, J. Asorey, S. Avila, O. Ballester, M. Banerji, W. Barkhouse *et al.*, Astrophys. J. Suppl. Ser. **239**, 18 (2018).
[26] C. P. Ahn, R. Alexandroff, C. Allende Prieto, S. F. Anderson, T. Anderton, B. H. Andrews, É. Aubourg, S. Bailey, E. Balbinot, R. Barnes *et al.*, Astrophys. J. Suppl. Ser. **203**, 21 (2012).
[27] J. C. Richardson, M. J. Irwin, A. W. McConnachie, N. F. Martin, A. L. Dotter, A. M. N. Ferguson, R. A. Ibata, S. C. Chapman, G. F. Lewis, N. R. Tanvir *et al.*, Astrophys. J. **732**, 76 (2011).
[28] A. W. McConnachie, Astron. J. **144**, 4 (2012).
[29] O. Newton, M. Cautun, A. Jenkins, C. S. Frenk, and J. C. Helly, Mon. Not. R. Astron. Soc. **479**, 2853 (2018).
[30] L. E. Strigari, J. S. Bullock, M. Kaplinghat, J. D. Simon, M. Geha, B. Willman, and M. G. Walker, Nature (London) **454**, 1096 (2008).
[31] J. Bland-Hawthorn and O. Gerhard, Annu. Rev. Astron. Astrophys. **54**, 529 (2016).
[32] G. Besla, arXiv:1511.03346.
[33] R. P. van der Marel, D. R. Alves, E. Hardy, and N. B. Suntzeff, Astron. J. **124**, 2639 (2002).
[34] G. Pietrzyński, D. Graczyk, W. Gieren, I. B. Thompson, B. Pilecki, A. Udalski, I. Soszyński, S. Kozłowski, P. Konorski, K. Suchomska *et al.*, Nature (London) **495**, 76 (2013).
[35] X. Luri, L. Chemin, G. Clementini, H. E. Delgado, P. J. McMillan, M. Romero-Gómez, E. Balbinot, A. Castro-Ginard, R. Mor*et al.* (Gaia Collaboration), Astron. Astrophys. **649**, A7 (2021).
[36] P. D. Dobbie, A. A. Cole, A. Subramaniam, and S. Keller, Mon. Not. R. Astron. Soc. **442**, 1663 (2014).
[37] V. Scowcroft, W. L. Freedman, B. F. Madore, A. Monson, S. E. Persson, J. Rich, M. Seibert, and J. R. Rigby, Astrophys. J. **816**, 49 (2016).
[38] Astropy Collaboration, T. P. Robitaille, E. J. Tollerud, P. Greenfield, M. Droettboom, E. Bray, T. Aldcroft, M. Davis, A. Ginsburg, A. M. Price-Whelan *et al.*, Astron. Astrophys. **558**, A33 (2013).
[39] Astropy Collaboration, A. M. Price-Whelan, B. M. Sipőcz, H. M. Günther, P. L. Lim, S. M. Crawford, S. Conseil, D. L. Shupe, M. W. Craig, N. Dencheva *et al.*, Astron. J. **156**, 123 (2018).
[40] R. P. van der Marel, M. Fardal, G. Besla, R. L. Beaton, S. T. Sohn, J. Anderson, T. Brown, and P. Guhathakurta, Astrophys. J. **753**, 8 (2012).
[41] E. Patel, G. Besla, and K. Mandel, Mon. Not. R. Astron. Soc. **468**, 3428 (2017).
[42] T. Tepper-García, J. Bland-Hawthorn, and D. Li, Mon. Not. R. Astron. Soc. **493**, 5636 (2020).
[43] R. P. van der Marel, M. A. Fardal, S. T. Sohn, E. Patel, G. Besla, A. del Pino, J. Sahlmann, and L. L. Watkins, Astrophys. J. **872**, 24 (2019).
[44] J. B. Salomon, R. A. Ibata, B. Famaey, N. F. Martin, and G. F. Lewis, Mon. Not. R. Astron. Soc. **456**, 4432 (2016).
[45] J. Sick, S. Courteau, J.-C. Cuillandre, J. Dalcanton, R. de Jong, M. McDonald, D. Simard, and R. B. Tully, in *Galaxy Masses as Constraints of Formation Models*, edited by M. Cappellari and S. Courteau (Cambridge University Press, Cambridge, England, 2015), Vol. 311, pp. 82–85.
[46] A. Tamm, E. Tempel, P. Tenjes, O. Tihhonova, and T. Tuvikene, Astron. Astrophys. **546**, A4 (2012).
[47] Y.-S. Li and S. D. M. White, Mon. Not. R. Astron. Soc. **384**, 1459 (2008).
[48] S. Phelps, A. Nusser, and V. Desjacques, Astrophys. J. **775**, 102 (2013).
[49] J. D. Diaz, S. E. Koposov, M. Irwin, V. Belokurov, and N. W. Evans, Mon. Not. R. Astron. Soc. **443**, 1688 (2014).
[50] J. Peñarrubia, Y.-Z. Ma, M. G. Walker, and A. McConnachie, Mon. Not. R. Astron. Soc. **443**, 2204 (2014).
[51] J. Peñarrubia, F. A. Gómez, G. Besla, D. Erkal, and Y.-Z. Ma, Mon. Not. R. Astron. Soc. **456**, L54 (2016).
[52] S. L. J. Gibbons, V. Belokurov, and N. W. Evans, Mon. Not. R. Astron. Soc. **445**, 3788 (2014).
[53] A. H. W. Küpper, E. Balbinot, A. Bonaca, K. V. Johnston, D. W. Hogg, P. Kroupa, and B. X. Santiago, Astrophys. J. **803**, 80 (2015).
[54] Y. Sofue, Publ. Astron. Soc. Jpn. **67**, 75 (2015).
[55] P. F. de Salas, K. Malhan, K. Freese, K. Hattori, and M. Valluri, J. Cosmol. Astropart. Phys. 10 (2019) 037.
[56] E. V. Karukes, M. Benito, F. Iocco, R. Trotta, and A. Geringer-Sameth, J. Cosmol. Astropart. Phys. 05 (2020) 033.
[57] M. Cautun, A. Benítez-Llambay, A. J. Deason, C. S. Frenk, A. Fattahi, F. A. Gómez, R. J. J. Grand, K. A. Oman, J. F. Navarro, and C. M. Simpson, Mon. Not. R. Astron. Soc. **494**, 4291 (2020).
[58] I. Ablimit, G. Zhao, C. Flynn, and S. A. Bird, Astrophys. J. Lett. **895**, L12 (2020).
[59] G. Monari, B. Famaey, I. Carrillo, T. Piffl, M. Steinmetz, R. F. G. Wyse, F. Anders, C. Chiappini, and K. Janßen, Astron. Astrophys. **616**, L9 (2018).
[60] A. J. Deason, A. Fattahi, V. Belokurov, N. W. Evans, R. J. J. Grand, F. Marinacci, and R. Pakmor, Mon. Not. R. Astron. Soc. **485**, 3514 (2019).
[61] A. J. Deason, D. Erkal, V. Belokurov, A. Fattahi, F. A. Gómez, R. J. J. Grand, R. Pakmor, X.-X. Xue, C. Liu, C. Yang *et al.*, Mon. Not. R. Astron. Soc. **501**, 5964 (2021).
[62] M. Boylan-Kolchin, J. S. Bullock, S. T. Sohn, G. Besla, and R. P. van der Marel, Astrophys. J. **768**, 140 (2013).
[63] R. E. González, A. V. Kravtsov, and N. Y. Gnedin, Astrophys. J. **770**, 96 (2013).
[64] M. Cautun, C. S. Frenk, R. van de Weygaert, W. A. Hellwing, and B. J. T. Jones, Mon. Not. R. Astron. Soc. **445**, 2049 (2014).
[65] E. Patel, G. Besla, K. Mandel, and S. T. Sohn, Astrophys. J. **857**, 78 (2018).
[66] T. M. Callingham, M. Cautun, A. J. Deason, C. S. Frenk, W. Wang, F. A. Gómez, R. J. J. Grand, F. Marinacci, and R. Pakmor, Mon. Not. R. Astron. Soc. **484**, 5453 (2019).
[67] Z.-Z. Li, Y.-Z. Qian, J. Han, T. S. Li, W. Wang, and Y. P. Jing, Astrophys. J. **894**, 10 (2020).

[68] T. K. Fritz, A. Di Cintio, G. Battaglia, C. Brook, and S. Taibi, Mon. Not. R. Astron. Soc. **494**, 5178 (2020).
[69] M. K. Rodriguez Wimberly, M. C. Cooper, D. C. Baxter, M. Boylan-Kolchin, J. S. Bullock, S. P. Fillingham, A. P. Ji, L. V. Sales, and J. D. Simon, Mon. Not. R. Astron. Soc. **513**, 4968 (2022).
[70] G. Eadie and M. Jurić, Astrophys. J. **875**, 159 (2019).
[71] L. L. Watkins, R. P. van der Marel, S. T. Sohn, and N. W. Evans, Astrophys. J. **873**, 118 (2019).
[72] W. Wang, J. Han, M. Cautun, Z. Li, and M. N. Ishigaki, Sci. China Phys. Mech. Astron. **63**, 109801 (2020).
[73] M. A. Fardal, M. D. Weinberg, A. Babul, M. J. Irwin, P. Guhathakurta, K. M. Gilbert, A. M. N. Ferguson, R. A. Ibata, G. F. Lewis, N. R. Tanvir *et al.*, Mon. Not. R. Astron. Soc. **434**, 2779 (2013).
[74] P. R. Kafle, S. Sharma, G. F. Lewis, A. S. G. Robotham, and S. P. Driver, Mon. Not. R. Astron. Soc. **475**, 4043 (2018).
[75] N. W. Evans and M. I. Wilkinson, Mon. Not. R. Astron. Soc. **316**, 929 (2000).
[76] P. Côté, M. Mateo, W. L. W. Sargent, and E. W. Olszewski, Astrophys. J. Lett. **537**, L91 (2000).
[77] L. L. Watkins, N. W. Evans, and J. H. An, Mon. Not. R. Astron. Soc. **406**, 264 (2010).
[78] E. J. Tollerud, R. L. Beaton, M. C. Geha, J. S. Bullock, P. Guhathakurta, J. S. Kalirai, S. R. Majewski, E. N. Kirby, K. M. Gilbert, B. Yniguez *et al.*, Astrophys. J. **752**, 45 (2012).
[79] J. Veljanoski, A. D. Mackey, A. M. N. Ferguson, A. P. Huxor, P. Côté, M. J. Irwin, N. R. Tanvir, J. Peñarrubia, E. J. Bernard, M. Fardal *et al.*, Mon. Not. R. Astron. Soc. **442**, 2929 (2014).
[80] A. Tamm, E. Tempel, P. Tenjes, O. Tihhonova, and T. Tuvikene, Astron. Astrophys. **546**, A4 (2012).
[81] R. E. Gonzalez, A. V. Kravtsov, and N. Y. Gnedin, Astrophys. J. **793**, 91 (2014).
[82] M. McLeod, N. Libeskind, O. Lahav, and Y. Hoffman, J. Cosmol. Astropart. Phys. 12 (2017) 034,
[83] E. Carlesi, Y. Hoffman, J. G. Sorce, and S. Gottlöber, Mon. Not. R. Astron. Soc. **465**, 4886 (2017).
[84] M. Boylan-Kolchin, G. Besla, and L. Hernquist, Mon. Not. R. Astron. Soc. **414**, 1560 (2011).
[85] G. Besla, N. Kallivayalil, L. Hernquist, B. Robertson, T. J. Cox, R. P. van der Marel, and C. Alcock, Astrophys. J. **668**, 949 (2007).
[86] M. Geha, R. H. Wechsler, Y.-Y. Mao, E. J. Tollerud, B. Weiner, R. Bernstein, B. Hoyle, S. Marchi, P. J. Marshall, R. Muñoz *et al.*, Astrophys. J. **847**, 4 (2017).
[87] S. Malhotra, D. N. Spergel, J. E. Rhoads, and J. Li, Astrophys. J. **473**, 687 (1996).
[88] T. C. Licquia, J. A. Newman, and M. A. Bershady, Astrophys. J. **833**, 220 (2016).
[89] E. Vasiliev and V. Belokurov, Mon. Not. R. Astron. Soc. **497**, 4162 (2020).
[90] E. N. Kirby, J. G. Cohen, P. Guhathakurta, L. Cheng, J. S. Bullock, and A. Gallazzi, Astrophys. J. **779**, 102 (2013).
[91] P. Villanueva-Domingo and F. Villaescusa-Navarro, Astrophys. J. **937**, 115 (2022).
[92] H. Shao, F. Villaescusa-Navarro, P. Villanueva-Domingo, R. Teyssier, L. H. Garrison, M. Gatti, D. Inman, Y. Ni, U. P. Steinwandel, M. Kulkarni *et al.*, Astrophys. J. **944**, 27 (2023).
[93] S. Anagnostidis, A. Thomsen, T. Kacprzak, T. Tröster, L. Biggio, A. Refregier, and T. Hofmann, arXiv:2211.12346.
[94] P. Villanueva-Domingo, Halographnet (2021), 10.5281/zenodo.5676528.
[95] https://github.com/PabloVD/HaloGraphNet.
[96] M. Fey and J. E. Lenssen, Fast graph representation learning with PyTorch geometric (2019), https://github.com/pyg-team/pytorch-geometric.
[97] https://www.camel-simulations.org.