Bayesian Cosmic Void Finding with Graph Flows

Leander Thiele [1 2]

Abstract

Cosmic voids contain higher-order cosmological information and are of interest for astroparticle physics. Finding genuine matter underdensities in sparse galaxy surveys is, however, an underconstrained problem. Traditional void finding algorithms produce deterministic void catalogs, neglecting the probabilistic nature of the problem. We present a method to sample from the stochastic mapping from galaxy catalogs to arbitrary void definitions. Our algorithm uses a deep graph neural network to evolve "test particles" according to a flow-matching objective. We demonstrate the method in a simplified example setting but outline steps to generalize it towards practically usable void finders. Trained on a deterministic teacher, the model performs well but has considerable stochasticity which we interpret as regularization. Cosmological information in the predicted void catalogs outperforms the teacher. On the one hand, our method can cheaply emulate existing void finders with apparently useful regularization. More importantly, it also allows us to find the Bayes-optimal mapping between observed galaxies and any void definition. This includes definitions operating at the level of simulated matter density and velocity fields.

1. Introduction

The late-time density field has grown gravitationally out of tiny initial perturbations into the intricate cosmic web. Extracting information about fundamental physics from this large-scale structure is one of the primary goals in current observational cosmology. Due to nonlinear growth, an optimal analysis method is difficult to conceive. Thus, various summary statistics have been developed to probe different aspects of the density field. Volume-wise the large-scale structure is dominated by underdense regions: cosmic voids. Voids have been recognized as holding unique cosmological information (see e.g., Pisani et al., 2019; Moresco et al., 2022; Schuster et al., 2023; 2024; 2025; Contarini et al., 2026). Finding voids has been an active area of research, an attempt at a comprehensive list of void finders being (El-Ad & Piran, 1997; Hoyle & Vogeley, 2002; Padilla et al., 2005; Neyrinck, 2008; Lavaux & Wandelt, 2010; Sutter et al., 2015; Elyiv et al., 2015; Ruiz et al., 2015; Zhao et al., 2016; Banerjee & Dalal, 2016; Nadathur et al., 2019; Ruiz et al., 2019; Douglass et al., 2022; Paz et al., 2023; Monllor-Berbegal et al., 2025; Ghafour et al., 2025; Sartori et al., 2026). It has been recognized that due to the multi-scale nature of the cosmic web the definition of a void is not unique, and indeed an "optimal" definition depends on the science case (Colberg et al., 2008; Cautun et al., 2018; Paillas et al., 2019). Additionally, void finders can be modified through postprocessing of the output catalogs (Ronconi & Marulli, 2017; Verza et al., 2025).

The void finding literature has strongly focused so far on addressing the problem with a deterministic approach, neglecting the fact that it naturally maps to probabilistic methods. Restricting the discussion to the case of galaxy redshift surveys, the observables are sparse and biased tracers of the density field. Thus, finding the underdense zones is an underconstrained problem. What seems more meaningful to consider is the probability distribution over void catalogs conditioned on galaxy catalogs

$$P(\{v_\alpha\}_{\alpha=1\ldots N_v}|\{g_\beta\}_{\beta=1\ldots N_g})\,, \tag{1}$$

where the v_α and g_β are tuples containing void and galaxy properties, respectively, and the catalog size N_v is a random variable. Traditional void finders adopt deterministic definitions of $\{v_\alpha\}$ such that this distribution collapses to a delta function. There are some algorithms computing confidence scores for individual voids, but due to void exclusion and similar effects such factorization of the distribution is a poor approximation: one needs to take correlations between voids into account.

Discarding the probabilistic character of the problem is not only problematic in a formal sense. It is tied to the observed instability of existing void finders (e.g., Sartori et al., 2026).

[1]Center for Data-Driven Discovery, Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan [2]Kavli IPMU (WPI), UTIAS, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8583, Japan. Correspondence to: Leander Thiele <leander.thiele@ipmu.jp>.

Proceedings of the AI4Physics Workshop at the 43rd International Conference on Machine Learning (AI4Physics@ICML 2026), Seoul, South Korea. 2026. Copyright 2026 by the author(s).

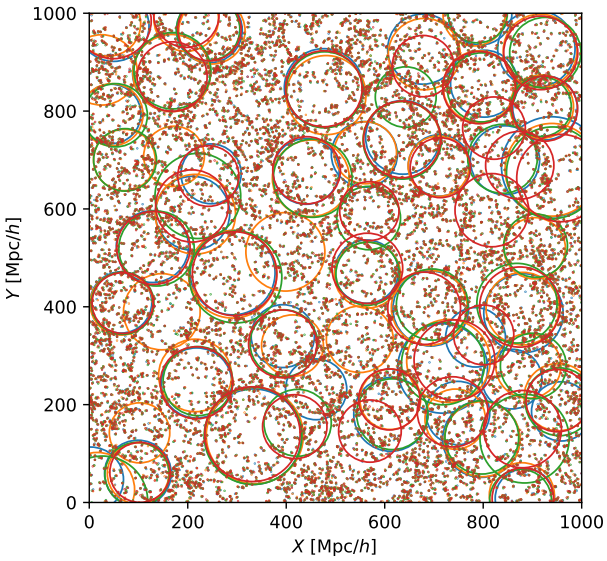

Figure 1. Illustration of the instability of deterministic void finders. Four times, the same galaxies are randomly jittered by $\sigma = 1\,h^{-1}$Mpc. The corresponding VIDE voids are shown as circles (matching in effective radius). There are substantial differences between the void catalogs.

We demonstrate this instability in Fig. 1 where VIDE void catalogs obtained after small random jitters of the underlying galaxies are displayed. We see that even though voids (here with minimum radius $60\,h^{-1}$Mpc) would be expected to decouple from small-scale perturbations, the resulting void catalogs are quite different. It has also been noticed that computing catalogs over random subsamples of a given galaxy field and averaging their statistics leads to significantly improved cosmological constraining power (Liu et al., 2026).

One approach towards Bayesian void finding as defined by Eq. (1) is through field-level reconstruction of the entire density field where the voids can then be defined (Leclercq et al., 2015b;a; Malandrino et al., 2026). While theoretically optimal, the technique is computationally expensive and has to sacrifice small-scale information to be feasible.

Another approach would be a local classification of the density field so as to match existing local algorithms (Sousbie, 2013; Cautun et al., 2013; Forero-Romero et al., 2009; Pomarède et al., 2017). Indeed, deep learning has been used to learn such classifications (Kumagai et al., 2025; Kololgi et al., 2025). We see value in catalogs of discrete voids as opposed to point-wise classifications, however. The field has developed many tools to analyze and theoretically describe such catalogs. Discrete voids are also convenient for cross-correlations, including with the CMB (Cai et al., 2017; Sartori et al., 2025; Granett et al., 2008; Melchior et al., 2014) and with gamma rays (Furniss et al., 2015; Arcari et al., 2022).

Thus, in this work we develop a general solution to computing samples (i.e. void catalogs) from the distribution Eq. (1). Our solution relies on deep learning, specifically graph neural networks trained with a flow matching objective. The training is performed in a supervised manner on cosmological simulations. We emphasize that our method is rather general in the sense that any void definition can be targeted. That is, the training data can be computed in any way from the available simulations, including from the initial or final density and velocity field. For any desired void definition motivated by a specific scientific question, our method provides the Bayes-optimal void finder.

After laying out the model design and training procedure in Sec. 2, we will present results through some summary statistics in Sec. 3. The discussion in Sec. 4 puts this work in the context of cosmological inference and observations in general machine learning. We conclude in Sec. 5.

2. Methods

Mapping the construction of a void catalog to a generative machine learning problem seems to require a somewhat non-standard setup. We solve the problem by using *test particles* whose coordinates and attributes correspond to the desired void positions and properties. Indeed, test particles are a known solution for image segmentation problems (e.g., Grady, 2006) and have been used for reconstructions of the cosmic web (Burchett et al., 2020). The presentation will be restricted to a cubic box at a fixed redshift with galaxies in real space. Future work will introduce the (tractable) complications necessary to deal with observational lightcones.

2.1. Flow Matching

The mentioned test particles are evolved by a neural network to track as closely as possible so-called interpolants. The method is known as flow matching (Lipman et al., 2023; Liu et al., 2022) which is a limit of stochastic interpolants (Albergo et al., 2023; Chen et al., 2024). Applications of similar methods in cosmology are found in Cuesta-Lazaro et al. (2024); Cuesta-Lazaro & Mishra-Sharma (2024); Horowitz et al. (2025); Kannan et al. (2025).

We introduce a fictitious time $t \in [0, 1]$ which parameterizes the flow between two distributions. At $t = 0$ the test particles follow a trivial base distribution, while at $t = 1$ they are supposed to follow the complicated distribution over void catalogs we aim to sample from.

The αth test particle is a tuple $v_\alpha(t)$ which contains the position of the particle and further attributes:

$$v_\alpha(t) = (\vec{r}_\alpha(t), \vec{a}_\alpha(t))\,. \tag{2}$$

For this proof-of-concept work, we take $\vec{a}_\alpha \in \mathbb{R}^1$ corresponding to the volume of the targeted cosmic void. There is no limitation on other void properties to predict as part of the $\vec{a}_\alpha$.

At $t = 0$ there are N_p test particles Poisson-initialized in the box at $v_\alpha(0)$. Time evolution of the test particles is defined as the ODE

$$\frac{dv_\alpha(t)}{dt} = f_\phi(v_\alpha(t), \{v_{\alpha'}(t)\}_{\alpha'=1\ldots N_p}, \{g_\beta\}_{\beta=1\ldots N_g}, t)\,, \tag{3}$$

where g_β describe the galaxies in which the voids are to be found. The graph neural network f_ϕ is described in more detail in Sec. 2.3.

During training, fictitious time is sampled uniformly $t \sim [0, 1]$ and optimization of the neural network is performed to minimize the objective

$$\int_0^1 dt \sum_{B \in \substack{\text{training}\\ \text{set}}} \sum_{\alpha=1\ldots N_p} \left| f_\phi(v_\alpha(t), \ldots) - \frac{d\hat{v}_\alpha(t)}{dt} \right|^2 , \tag{4}$$

where B is a simulated box and we introduced the *interpolants*

$$\hat{v}_\alpha(t) = (1-t)v_\alpha(0) + t\hat{v}_\alpha(1)\,. \tag{5}$$

Here, $\hat{v}_\alpha(1)$ corresponds to an individual void contained in the training data. When trying to match test particles to cosmic voids, two problems arise.

First, we do not know the number of voids N_v a priori, so a one-to-one matching between test particles and voids is impossible. We solve this by oversampling, i.e. using $N_p > N_v$ test particles. Most voids in a training sample get assigned multiple interpolant indices α such that the total sums to N_p. We randomly set the number of endpoints per void such that in expectation it is proportional to the void's volume and has the minimum variance around it.

Second, given that a typical training sample (simulation box) in our setup contains hundreds of voids, *which* test particle α should be assigned to a given void? In the original flow matching framework, the assignment would be random. Such a choice of interpolants would be very difficult to learn, however, since the test particles would need to traverse large distances across the box and this would be unnatural in our locality-biased graph neural network. We perform the pairing by minimizing the sum of squared Euclidean distances between the initial random test particle positions $\vec{r}_\alpha(0)$ and the void catalog, using the Hungarian algorithm (Kuhn, 1955). To avoid degeneracy in the Hungarian assignment, we Gaussian "jitter" the void positions in $\hat{v}_\alpha(1)$ with $0.2\,h^{-1}$Mpc standard deviation.

The resulting construction of interpolants is in the spirit of "data-dependent couplings" (Albergo et al., 2024) but differs from previous literature in that it acts as a permutation, not a continuous transformation. We point out that the construction might superficially resemble minibatch optimal transport (Pooladian et al., 2023; Tong et al., 2024) but is very different due to the couplings between test particles (the existence of minibatch-OT as the zero-coupling limit provides a useful check, however). Note that the possibility to minimize total traversed path length does not exist in diffusion models.

During inference, we integrate Eq. (3) according to the Euler rule

$$v_\alpha(t_{i+1}) = v_\alpha(t_i) + (t_{i+1} - t_i) f_\phi(v_\alpha(t_i), \ldots) \tag{6}$$

over the time interval $t \in [0, 1]$. We find that the resulting paths show somewhat more curvature at the beginning and end, thus opt for the stepping scheme

$$t_i = 0.9(3\tau_i^2 - 2\tau_i^3)^{1.5} + 0.1\tau_i\,, m\tau_i = 0 \ldots m-1 \tag{7}$$

where we take $m = 32$ integration steps. Our results are not very sensitive to this particular choice of time steps t_i. In Fig. 11 in the appendix we demonstrate that the integrated trajectories are very well converged.

As we oversample $N_p > N_v$, the integration of $v_\alpha(1)$ needs to be followed by a procedure to combine test particles into unique voids. We do this clustering using the friends-of-friends (FoF) algorithm (Huchra & Geller, 1982; Davis et al., 1985) on the $\vec{r}$-part of the prediction. Void properties are then computed as the mean of the grouped $v_\alpha(1)$. The FoF linking length is a hyperparameter but we find little dependence on it for reasonable values. Thanks to recent developments on differentiable clustering algorithms (Horowitz & Bayer, 2025) our entire void finder could be made differentiable. Differentiable pipelines from initial conditions to summary statistics can enable more efficient parameter estimation, for example with Hamiltonian Monte Carlo or raytracing (Behroozi, 2025) sampling.

2.2. Training data

For the purposes of this initial exposition of the deep learning method, we simplify the learning problem to its essential properties. Thus, we use Quijote simulations (Villaescusa-Navarro et al., 2020) of side length $1\,h^{-1}$Gpc run with 512^3 particles evolving under gravity only. Galaxies are inserted into the simulations according to a halo-occupation-distribution (HOD) prescription described in Sec. 2.2.1. We work at redshift $z = 0$ in real space.

For training, we use 12000 simulations from the BSQ set in which 5-parameter ΛCDM is varied over a wide prior (Bairagi et al., 2025). Another 1000 BSQ simulations are used for validation and testing, each. Each of the BSQ simulations is populated with 8.6 randomly drawn HODs on average. The resulting 115707 galaxy catalogs

$\log M_0/M_{\text{min}}$	$[0, 1]$
$\log M_1/M_0$	$[0, 3]$
$\sigma_{\log M}$	$[0.1, 0.8]$
α	$[0.2, 1.5]$

Table 1. Prior bounds for the HOD parameters.

are further augmented through the usual 48 rotations and transpositions.

In addition, we use 15000 simulations at the fiducial point populated with the fiducial HOD parameters for evaluation.

2.2.1. Galaxies

We use a standard HOD (Berlind & Weinberg, 2002; Cooray & Sheth, 2002; Wechsler & Tinker, 2018) to populate the gravity-only simulations with galaxies. The form is taken from Zheng et al. (2005) and explicitly written out in Appendix B. HOD parameters are sampled according to uniform priors with bounds given in Table 1. The HOD fiducial point is taken as the mean of the priors.

Using M_{min}, we fix the overall number of galaxies per training sample to $N_g = 10^5$, corresponding to a number density of $n = 10^{-4}(\text{Mpc}/h)^{-3}$. Given the large parameter variations we believe it makes physical sense to constrain the number density. Furthermore, for this initial work it has computational advantages as each training sample demands about equal computation. At the Quijote fiducial cosmology and HOD fiducial point, the corresponding $\log M_{\text{min}}/h^{-1}M_\odot = 13.78$. We are working with a rather sparse galaxy sample for simplicity in this initial demonstration but do not see any fundamental limitation on galaxy density.[1]

2.2.2. Void definitions

The primary void definition we use in this work are VIDE-voids in the galaxy field (Sutter et al., 2015). At first glance, this seems paradoxical: this void definition uses the same information as is available during inference of our graph net flow. Thus, we are seemingly not taking advantage of our algorithm's Bayesian character at all and simply aiming to replicate a deterministic void finder. As we shall see, the strong inductive bias in our architecture and training objective lets the network converge to a non-deterministic void finder. We will argue in Sec. 4 that in some sense the trained network can be seen as *outperforming* its VIDE teacher.

Regardless of these points, it is also beneficial for this initial work to aim for a well-understood baseline performance. In future work we will learn to predict voids defined on the simulated density and velocity fields and thus take full advantage of our Bayes-optimal method.

The VIDE-catalogs from the galaxy field are truncated at a minimum void radius of $R_{\text{min}} = 60\,h^{-1}$Mpc. This value is appropriate for the low tracer density we are working with (mean galaxy separation is $\sim 21.5\,h^{-1}$Mpc). Given the range of cosmologies and HODs in our training data, individual boxes contain $N_v \sim 100 \ldots 600$ voids after this cut. In the following, we denote them *galaxy voids*.

As a test, we also use void catalogs based on the dark matter field. These voids are found using VIDE run directly on the simulation particles. We choose this setup so as to test the limits of our method since galaxy voids have little in common with these dark matter-defined ones. For computational reasons we subsample the simulation particles to $10\,\%$. Due to the very different tracer density, we cut these catalogs at $R_{\text{min}} = 25.04\,h^{-1}$Mpc so as to obtain equal void numbers with the galaxy voids at the fiducial point. We will see that voids constructed from these high-density tracers are very difficult to identify from the galaxies, introducing large stochasticity and complicating the learning problem. In the following, we denote them *DM voids*.

We always use VIDE's default volume-weighted barycenter definition for the positions and note that this is a highly nonlocal definition depending on all Voronoi cells belonging to a given void. One might expect void position definitions based on local density minima to be easier to learn. Since we are interested to explore the capabilities of our proposed algorithm, we choose the more difficult definition.

2.3. Architecture and Training

We construct the test particle fictitious velocity $f_\phi(v_\alpha(t), \{v_{\alpha'}(t)\}_{\alpha'=1\ldots N_p}, \{g_\beta\}_{\beta=1\ldots N_g}, t)$ as a graph neural network (Gori et al., 2005; Scarselli et al., 2009). There are two types of nodes, namely test particles P and galaxies G. We also include edges PP and PG. Appendix C describes how we compute the input node and edge features. We emphasize that the graph topology depends on fictitious time.

The network only updates the test particle features x^P_α, since the computational cost of updating the galaxy features x^G_β and edges $e^{PG}_{\alpha\beta}$, $e^{PP}_{\alpha\alpha'}$ would be prohibitive. The network updates x^P_α using message passing (Gilmer et al., 2017; Battaglia et al., 2018; Hamilton et al., 2018). Messages from galaxies to particles are computed in the lth round of message passing as

$$m^{GP,l}_\alpha = \text{Aggr}_{\beta\in\text{knn}(\alpha)} g_l(\{\tilde{x}^P_\alpha, x^G_\beta, e^{GP}_{\alpha\beta}\})\,, \qquad (8)$$

where Aggr is a permutation-invariant aggregation over the k-nearest neighbors (in 3-dimensional Euclidean space) and

[1] Compare with $n \sim 10^{-3}(\text{Mpc}/h)^{-3}$ for the final DESI LRG sample. We are also constrained by the 32GB of VRAM on the V100s we have available.

g_l is a multi-layer perceptron (MLP). Analogous messages m^{PP} are computed from the neighboring test particles. Although the test particle features x_α^P are updated through subsequent rounds of message passing, we always concatenate them with the initial features to give $\tilde{x}_\alpha^P$. The motivation for such skip-connections is to prevent the oversmoothing which can occur in deep neural networks. We take the aggregation functions as mean and maximum in the ratio 3:1 between different message passing rounds.

Test particle features are updated according to the rule

$$x_\alpha^P \leftarrow \text{LayerNorm}(x_\alpha^P + \eta \text{FiLM}_t(h_l(\{m_\alpha^{PP,l}, m_\alpha^{GP,l}\}))) \,, \tag{9}$$

with layer normalization (Ba et al., 2016) for stability, feature-wise linear modulation (Perez et al., 2018) to include conditioning on the fictitious flow time t, η a learned feature-wise scaling, and h_l another MLP.

Our network architecture does not respect the periodic boundary conditions of the simulation boxes. This choice is made deliberately looking towards the planned application of the method to observational lightcones. In order to reduce boundary artifacts, we ignore voids closer than $20\,h^{-1}$Mpc to the box edges. We also do not construct the architecture in a rotationally or translationally equivariant way, due to the computational cost of respecting these relatively small symmetries.

We use the following numerical settings: $k_{GP} = 128$, $k_{PP} = 16$ for the k-nearest neighbors, MLP width 256 and depth 2, 24 rounds of message passing, for a total of 22 million network parameters. Activation functions are exponential linear units (ELU, Clevert et al., 2016).

The scaling of our architecture is naively $O(\lambda N_p^2 + N_p N_g)$ where λ depends on hyperparameters. With the choices adopted in this work the second term dominates. The given scaling expression assumes a constant survey volume with receptive fields fixed in length units. It is reasonable to assume that with larger galaxy density there would be more interest in smaller voids and thus the appropriate receptive field could be smaller. Therefore, the practical scaling might be better than the naive expectation.

We implement the network in PyTorch (Paszke et al., 2019) using the PyG library[2] and optimize it with the Adam algorithm (Kingma & Ba, 2017). Some loss curves are shown in Fig. 2. The long training runs correspond to the final models examined more closely in Sec. 3. For both galaxy voids (orange) and DM voids (green) we observe typical convergence, albeit to rather different values due to the much larger stochasticity in the latter case. The galaxy voids case may benefit from further training beyond our computational budget.

[2] https://pytorch-geometric.readthedocs.io

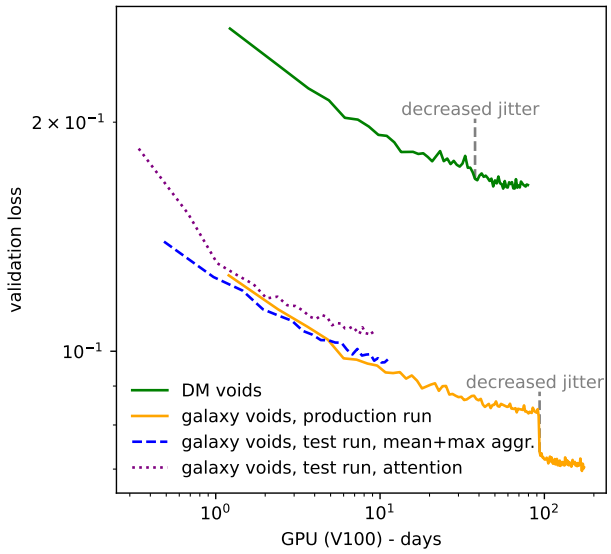

Figure 2. Validation loss curves for final production runs (orange, green) and illustrative trial runs with smaller models (blue, purple). At the indicated points, the random jittering of target void positions was decreased to the final value.

Amongst many trial runs, we point out a comparison of two which might be interesting for practitioners. The blue dashed curve corresponds to a network with similar structure to our final model but smaller scale for faster exploration. The purple dotted curve is for a model of similar size but with a multi-head attention mechanism (Veličković et al., 2018) instead of the traditional message aggregation. It is only computed over the k-nearest neighbors regions. We observe that the traditional aggregation strategy outperforms attention. We believe that the advantage of traditional architectures over transformer-inspired ones is a rather general phenomenon for cosmological data, the reason being the smoothness of typical fields. From theoretical works, it is known that transformers exploit sparsity (e.g., Mousavi-Hosseini et al., 2025). Indeed, previous works found analogous advantages of traditional CNN and GNN aggregations over attention (Hwang et al., 2023; Kakadia et al., 2025), although the opposite has been found as well (Kololgi et al., 2025). We thus caution against extrapolating results from natural language or everyday images to the rather different problems arising in cosmology.

3. Results

In this section, we begin by considering the case of galaxy voids. We remind the reader that the desired mapping is, in principle, deterministic and perfectly learnable. We first present results computed at the fiducial point (of both cosmology and HOD) and defer parameter variations to the discussion in Sec. 4. We use $N_p = 10^3$ test particles and

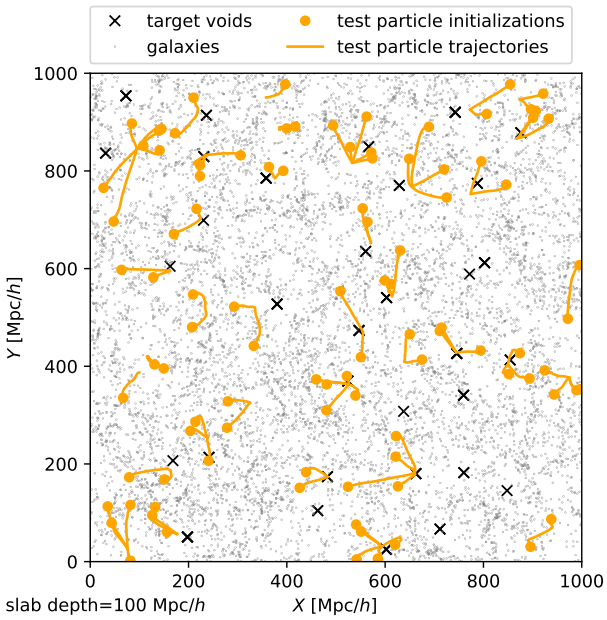

Figure 3. Slab of $100\,h^{-1}$Mpc thickness in a test sample. Grey points are galaxies and black crosses ground-truth VIDE voids. Test particles are randomly initialized in the filled orange circles and then evolved by the graph network along the orange trajectories. The test particles also carry a prediction of void volume which is not visualized. See Fig. 10 in the appendix for multiple random initializations of the test particles.

adopt a linking length $b = 15\,h^{-1}$Mpc for the FoF step (recall the void size cut $R_{\text{min}} = 60\,h^{-1}$Mpc).

In Fig. 3 we begin by presenting a visual impression of the trained void finder's performance. We show a slab of $100\,h^{-1}$Mpc thickness in one of the test samples. Galaxies are shown as grey points, the target void centers (computed by VIDE) as black crosses, and the test particle trajectories as orange lines. The randomly sampled starting points of these trajectories are marked as filled orange circles. The trajectories have small curvatures. Further, whenever multiple test particles converge because the model deems them to belong to a single void, the convergence is very good. These observations indicate a well-trained model. In many instances there is a clear correspondence between trajectory ends and ground-truth voids. This is, however, not universally the case. Thus, the trained void finder behaves differently than the mapping it was trained on.

A useful check is how many test particles end up being grouped together in voids of various sizes. According to our construction of the training set, we expect the number of test particles converging towards a single void to be proportional to the void volume. In Fig. 4, we show how the model behaves relative to this expectation. For almost all void sizes the empirical behavior is correct, except for the smallest voids for which a $\sim 10\,\%$ discrepancy appears.

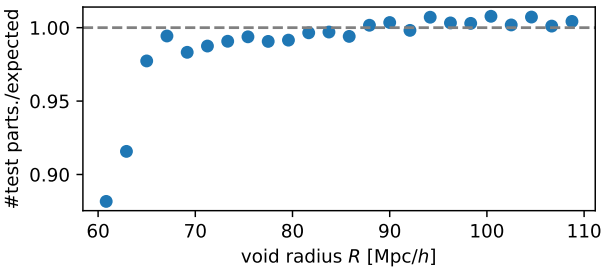

Figure 4. Check for the consistency condition that, in expectation, number of converging test particles should be proportional to void volume.

Our interpretation is that for a small set of "failed" test particles, the model is unable to find a position towards which it should evolve them. The adopted behavior in these cases seems to be to evolve the position only very little and to drive the volume prediction towards the lower end of the prior. Due to the decreasing void size function, this could be Bayes-optimal given the model's limitations.

In Fig. 5 we compare the main summary statistics between model predictions (orange) and ground-truth VIDE-voids (black).

The void size function (VSF) in the top panel matches relatively well except for an overprediction at small radii and a slight underprediction at large radii. The overprediction is consistent with the previously developed picture of "failed" test particles. The underprediction can likely be understood as convergence to the mean. For the void size function, we also show the statistical scatter as error bars. Light orange is the scatter including model stochasticity (i.e. the model only evolves a single population of test particles per test sample), while dark orange is the sample variance alone (i.e. the model is averaged over multiple draws of test particles). We observe that the trained model has substantial stochasticity but ends up matching the scatter in the VIDE VSF quite closely. When the model's stochasticity is averaged over, however, the VSF is substantially *less* noisy than its teacher. We will discuss this more in Sec. 4.

The void-galaxy correlation function (i.e. average void profile) and void-void correlation function[3] in the bottom panel are consistent with our previously developed picture of a well-trained model with a small population of "failed" test particles. Both statistics match the ground truth relatively well while indicating slightly less underdense and slightly less mutually excluding voids. In the void-void correlation function we can clearly see a feature matching the FoF linking length $b = 15\,h^{-1}$Mpc.

A generative model can sometimes reproduce summary

[3]Correlation functions are computed with pycorr, https://py2pcf.readthedocs.io.

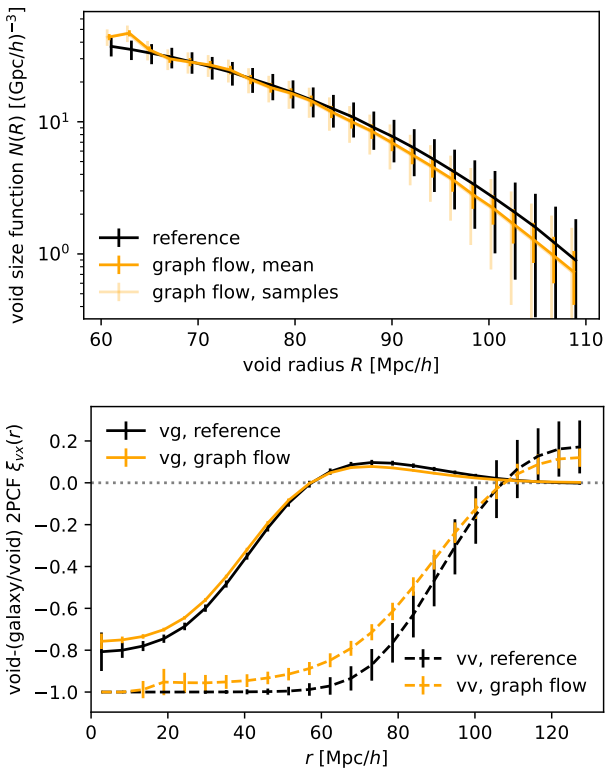

Figure 5. Summary statistics of galaxy voids at the fiducial cosmology and HOD. Black is VIDE ground truth, orange our model's prediction. *Top panel:* void size function. *Bottom panel:* void-galaxy and void-void correlation functions.

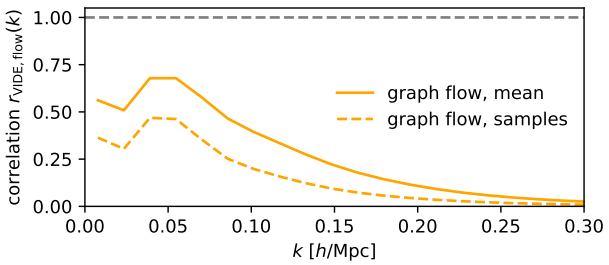

Figure 6. Cross-correlation between our model's output void catalog and the VIDE ground truth. Note that we have not subtracted any shot noise from the auto power spectra (it is not Poissonian). Thus, we are underestimating $r(k)$.

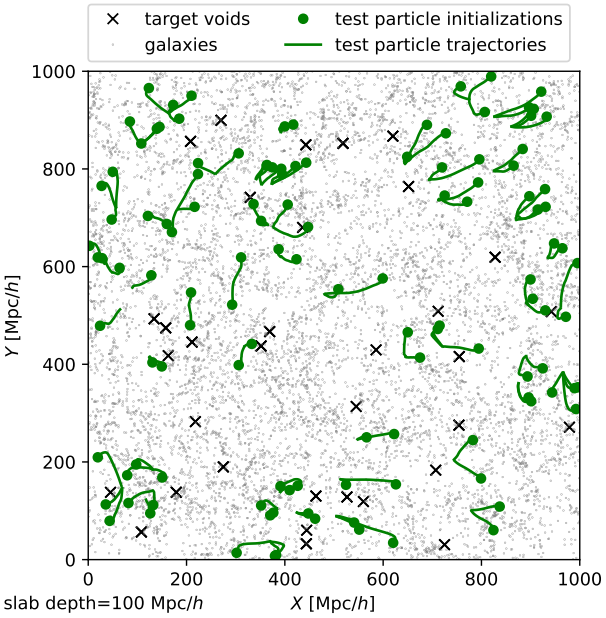

Figure 7. As Fig. 3 but for DM voids.

statistics quite well without actually solving the intended task. Therefore, in Fig. 6 we show the cross-correlation between our generated void catalogs and the reference ones. Note that as macroscopic objects voids have unknown non-Poissonian stochasticity and thus we cannot subtract shot noise from the auto power spectra. Therefore, $r(k)$ is systematically underestimated. The mean of the generated field yields higher $r(k)$ than the samples, consistent with lower stochasticity. The correlation decreases beyond $k \sim 0.05\,h\mathrm{Mpc}^{-1}$, roughly corresponding to the considered void size of $R > 60\,h^{-1}\mathrm{Mpc}$. While difficult to interpret, we conclude that our generated void catalogs are generally correlated with the reference ones, in agreement with the qualitative observations in Fig. 3.

3.1. DM voids

After the main results on galaxy voids, we now consider the more difficult problem of finding voids as identified by VIDE in the (subsampled) simulation particle field. Since for these DM voids we choose the size cut $R_{\mathrm{min}} = 25.04\,h^{-1}\mathrm{Mpc}$, we adopt a linking length $b = 5\,h^{-1}\mathrm{Mpc}$.

Fig. 7 shows example trajectories generated by the trained model, analogous to Fig. 3 for the galaxy voids case. We observe little correspondence between the trajectory endpoints and the ground truth voids. Trajectories are generally more curved as compared to the galaxy voids model. These observations agree with the intuition that the mapping between galaxies and DM voids has large stochasticity.

More quantitatively, the top panel in Fig. 8 shows the DM

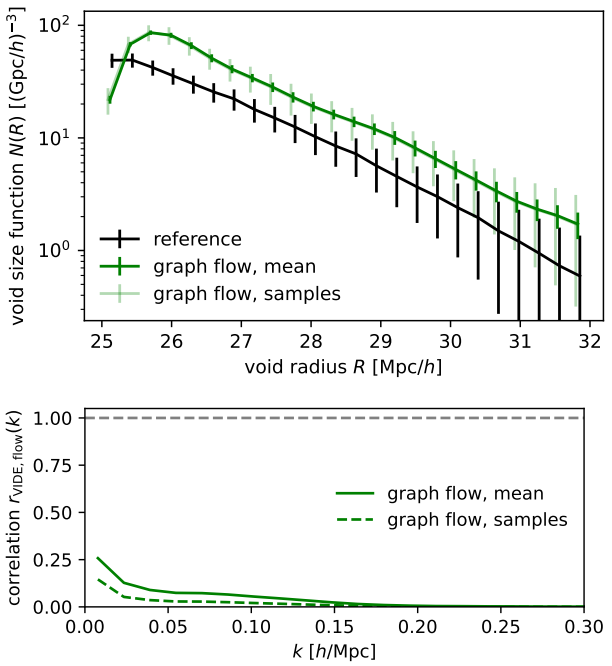

Figure 8. Top panel: Comparison of void size function for DM voids (c.f. Fig. 5). *Bottom panel:* Cross-correlation for DM voids (c.f. Fig. 6).

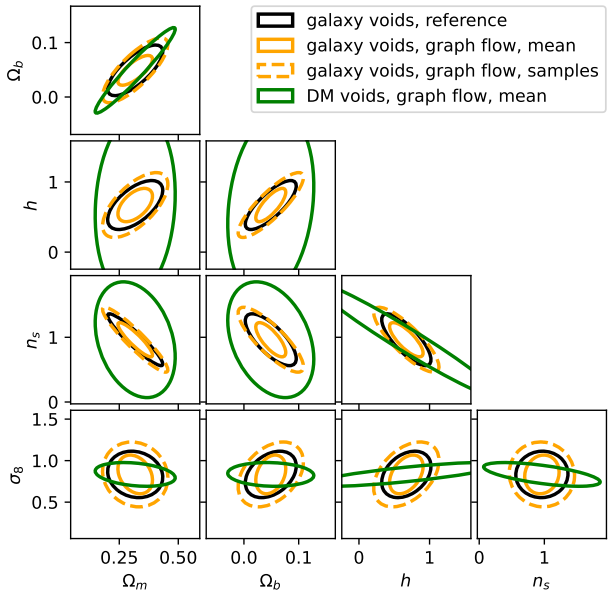

Figure 9. Visualization of the cosmological information in void size function and void-galaxy correlation function. Marginalization over HOD parameters is included.

VSF. The model generally overpredicts the VSF except for the smallest voids. This seems connected to test particles not converging as much as in the galaxy voids case. The picture is consistent with our previous observation that faced with large uncertainty the model produces a population of "failed" trajectories that do not converge with other particles. The bottom panel in Fig. 8 shows that the cross-correlation with VIDE's output is expectedly lower than in the case of galaxy voids, but importantly there is still non-zero correlation which means that the model has learnt something meaningful instead of just randomly reproducing summary statistics.

It is well-known that the void catalogs returned by traditional algorithms are highly dependent on tracer density (Schmidt et al., 2001; Sutter et al., 2014; Schuster et al., 2023). Thus, it is not surprising that supplied only with sparse galaxy catalogs our algorithm struggles to find the DM voids. Our results are consistent with the hypothesis that the network learned the Bayes-optimal mapping under limited information.

4. Discussion

After developing intuition for the kind of void finding algorithm our models have learned, we now turn to discuss usability and interpretation. A typical use for a void catalog would be to constrain cosmological parameters using summary statistics. We consider the combined data vector of void size function and void-galaxy correlation function. (The void-void correlation function has very low signal-to-noise.) The Fisher information, after marginalizing over HOD parameters, is shown in Fig. 9.[4] We emphasize that the figure is to be understood not as a forecast but rather as a convenient visualization of *local* signal-to-noise.

Focusing on the galaxy voids (black and orange contours) first, we observe that the orientation of ellipses is quite consistent between the VIDE result and our graph flow. Furthermore, the information from the mean of summary statistics in graph flow catalogs is higher than the one from VIDE's deterministic catalogs. Thus, the model's outputs respond correctly to cosmological parameters; no significant convergence towards the training set's mean is observed.

We also show the result from DM voids as found by our graph flow in green. Direct comparison of contours is difficult but we believe that by cutting catalogs to the same size at the fiducial point (using R_{min}) it is still meaningful. Statistics of DM voids are competitive only for σ_8. We thus conclude that more work is needed to find a void definition that is learnable by our graph flow and can substantially improve cosmological constraining power.

[4]Numerical derivatives are computed using the 1000 Quijote BSQ simulations we set aside for testing. The derivatives are approximated by linearizing the entire parameter space. Restricting to a smaller ball around the fiducial point yields some changes in the Fisher matrix but a consistent picture.

Going back to the galaxy voids, how do we interpret the fact that our graph flow has learned a non-deterministic void finder which in cosmological constraining power *outperforms* its teacher? Any finite-size model trained with finite resources will be biased away from the deterministic point defined by the VIDE teacher. The training algorithm introduces further biasing towards minima with good generalization properties. We therefore interpret our trained graph flow as a *regularized VIDE*. The exact regularization is of course impossible to write down but it seems to be useful as indicated by the cosmological information. Indeed, similar cases of a regularized student outperforming its teacher are known in the machine learning literature (Xie et al., 2020; Furlanello et al., 2018).

5. Conclusions

We have developed a Bayes-optimal method to probabilistically sample void catalogs conditioned on observed galaxies. The targeted void definition is arbitrary and therefore can rely on information that is only available in simulations. Our algorithm relies on a deep graph neural network that predicts fictitious velocities of test particles to match a defined flow. Through summary statistics and cosmological information we have demonstrated that our learned flow does indeed predict useful void catalogs.

Having developed the initial framework in an idealized situation (simulation box, real space, low tracer density, etc.), we will incorporate practical complications in future work. We aim to exploit the method for a practically usable void finder. A central question will be which target void definition should be adopted. We found that naively targeting VIDE voids in the simulation particles yields a mapping dominated by stochasticity. The DM voids our network is able to identify appear to contain less cosmological information than galaxy voids. Thus, we plan to identify "middle ground" definitions which can take advantage of our Bayesian method while still having high mutual information with sparse observed galaxies. We see promise in definitions that use the simulated velocity field.

On the machine learning side, the simple flow matching setup works well but could be improved by adding stochastic terms or replacing the ODE with a diffusive SDE. There might be room for reinforcement learning to improve the output void catalogs beyond the simple FoF clustering step we have been using. It will also be interesting to explore more uniform base densities, for example glassy initial distributions for the test particles instead of the current Poisson one.

The method of test particles evolving under a conditioned flow is much more general than void finding. Application to filament finding is underway. Another useful application could be insertion of halos or galaxies into low-resolution cosmological simulations, similar to, e.g., (Pandey et al., 2024; 2025a;b).

Extracting *robust* information from cosmological data is more difficult than signal-to-noise maximization. Indeed, direct application of deep learning to field-level data often leads to severe biases even in simple situations (e.g., Bayer et al., 2025). We believe that guard-railing the extracted information with understood physical concepts is a promising avenue towards integrating the power of machine learning into cosmological inference.

Acknowledgements

I thank Alice Pisani, Pierre Boccard, and Ben Horowitz for reading an earlier version of this manuscript and giving extremely helpful comments. I thank the organizers and participants of the Voids@CPPM 2025 workshop, in particular Alice Pisani, Nico Schuster, and Pierre Boccard. I thank Yu Liu and Cheng Zhao for sharing their work and discussions. I thank Carolina Cuesta-Lazaro, Giovanni Verza, Peter Behroozi, Xavier Prochaska, and Jia Liu for useful discussions. I thank the anonymous referee for constructive comments which helped improve the manuscript. LT is supported by JSPS under KAKENHI 24K22878 and 26K17136 and by the Royal Society under ICA\R2\252140. The Kavli IPMU is supported by World Premier International Research Center Initiative (WPI), MEXT, Japan.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

References

Albergo, M. S., Boffi, N. M., and Vanden-Eijnden, E. Stochastic Interpolants: A Unifying Framework for Flows and Diffusions. *arXiv e-prints*, art. arXiv:2303.08797, March 2023. doi: 10.48550/arXiv.2303.08797.

Albergo, M. S., Goldstein, M., Boffi, N. M., Ranganath, R., and Vanden-Eijnden, E. Stochastic interpolants with data-dependent couplings. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Arcari, S., Pinetti, E., and Fornengo, N. Got plenty of nothing: cosmic voids as a probe of particle dark matter. *J. Cosmol. Astropart. Phys.*, 2022(11):011, November 2022. doi: 10.1088/1475-7516/2022/11/011.

Ba, J. L., Kiros, J. R., and Hinton, G. E. Layer normaliza-

tion. 2016. URL https://arxiv.org/abs/1607.06450.

Bairagi, A., Wandelt, B., and Villaescusa-Navarro, F. The BIG SOBOL SEQUENCE: How many simulations do we need for simulation-based inference in cosmology? *Astron. Astrophys.*, 703:A301, November 2025. doi: 10.1051/0004-6361/202554602.

Banerjee, A. and Dalal, N. Simulating nonlinear cosmological structure formation with massive neutrinos. *J. Cosmol. Astropart. Phys.*, 2016(11):015, November 2016. doi: 10.1088/1475-7516/2016/11/015.

Battaglia, P. W. et al. Relational inductive biases, deep learning, and graph networks. 2018. URL https://arxiv.org/abs/1806.01261.

Bayer, A. E., Villaescusa-Navarro, F., Sharief, S., Teyssier, R., Garrison, L. H., Perreault-Levasseur, L., Bryan, G. L., Gatti, M., and Visbal, E. Field-level Comparison and Robustness Analysis of Cosmological N-body Simulations. *Astrophys. J.*, 989(2):207, August 2025. doi: 10.3847/1538-4357/adef4e.

Behroozi, P. The Ray Tracing Sampler: Bayesian Sampling of Neural Networks for Everyone. *arXiv e-prints*, art. arXiv:2510.25824, October 2025. doi: 10.48550/arXiv.2510.25824.

Berlind, A. A. and Weinberg, D. H. The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass. *Astrophys. J.*, 575 (2):587–616, August 2002. doi: 10.1086/341469.

Burchett, J. N., Elek, O., Tejos, N., Prochaska, J. X., Tripp, T. M., Bordoloi, R., and Forbes, A. G. Revealing the Dark Threads of the Cosmic Web. *Astrophys. J. Lett.*, 891 (2):L35, March 2020. doi: 10.3847/2041-8213/ab700c.

Cai, Y.-C., Neyrinck, M., Mao, Q., Peacock, J. A., Szapudi, I., and Berlind, A. A. The lensing and temperature imprints of voids on the cosmic microwave background. *Mon. Not. Roy. Astron. Soc.*, 466(3):3364–3375, April 2017. doi: 10.1093/mnras/stw3299.

Cautun, M., van de Weygaert, R., and Jones, B. J. T. NEXUS: tracing the cosmic web connection. *Mon. Not. Roy. Astron. Soc.*, 429(2):1286–1308, February 2013. doi: 10.1093/mnras/sts416.

Cautun, M., Paillas, E., Cai, Y.-C., Bose, S., Armijo, J., Li, B., and Padilla, N. The Santiago-Harvard-Edinburgh-Durham void comparison - I. SHEDding light on chameleon gravity tests. *Mon. Not. Roy. Astron. Soc.*, 476(3):3195–3217, May 2018. doi: 10.1093/mnras/sty463.

Chen, Y., Goldstein, M., Hua, M., Albergo, M. S., Boffi, N. M., and Vanden-Eijnden, E. Probabilistic Forecasting with Stochastic Interpolants and Föllmer Processes. *arXiv e-prints*, art. arXiv:2403.13724, March 2024. doi: 10.48550/arXiv.2403.13724.

Clevert, D.-A., Unterthiner, T., and Hochreiter, S. Fast and accurate deep network learning by exponential linear units (elus). 2016. URL https://arxiv.org/abs/1511.07289.

Colberg, J. M. et al. The Aspen-Amsterdam void finder comparison project. *Mon. Not. Roy. Astron. Soc.*, 387 (2):933–944, June 2008. doi: 10.1111/j.1365-2966.2008.13307.x.

Contarini, S., Verza, G., and Pisani, A. The era of precision cosmology with voids. 2026. URL https://arxiv.org/abs/2601.14362.

Cooray, A. and Sheth, R. Halo models of large scale structure. *Phys. Rep.*, 372(1):1–129, December 2002. doi: 10.1016/S0370-1573(02)00276-4.

Cuesta-Lazaro, C. and Mishra-Sharma, S. Point cloud approach to generative modeling for galaxy surveys at the field level. *Phys. Rev. D*, 109: 123531, Jun 2024. doi: 10.1103/PhysRevD.109.123531. URL https://link.aps.org/doi/10.1103/PhysRevD.109.123531.

Cuesta-Lazaro, C., Bayer, A. E., Albergo, M. S., Mishra-Sharma, S., Modi, C., and Eisenstein, D. J. Joint cosmological parameter inference and initial condition reconstruction with stochastic interpolants. In *NeurIPS 2024 Workshop: Machine Learning and the Physical Sciences*, 2024.

Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. Evolution of large-scale structure in a universe dominated by cold dark matter. *Astrophys. J.; (United States)*, 292:2, 05 1985. ISSN ISSN ASJOA. doi: 10.1086/163168. URL https://www.osti.gov/biblio/5014740.

Douglass, K., Veyrat, D., O'Neill, S., BenZvi, S., Zaidouni, F., and Guzzetti, M. VAST: the Void Analysis Software Toolkit. *The Journal of Open Source Software*, 7(77): 4033, September 2022. doi: 10.21105/joss.04033.

Duffy, A. R., Schaye, J., Kay, S. T., and Dalla Vecchia, C. Dark matter halo concentrations in the Wilkinson Microwave Anisotropy Probe year 5 cosmology. *Mon. Not. Roy. Astron. Soc.*, 390(1):L64–L68, October 2008. doi: 10.1111/j.1745-3933.2008.00537.x.

El-Ad, H. and Piran, T. Voids in the Large-Scale Structure. *Astrophys. J.*, 491(2):421–435, December 1997. doi: 10.1086/304973.

Elyiv, A., Marulli, F., Pollina, G., Baldi, M., Branchini, E., Cimatti, A., and Moscardini, L. Cosmic voids detection without density measurements. *Mon. Not. Roy. Astron. Soc.*, 448(1):642–653, March 2015. doi: 10.1093/mnras/stv043.

Forero-Romero, J. E., Hoffman, Y., Gottlöber, S., Klypin, A., and Yepes, G. A dynamical classification of the cosmic web. *Mon. Not. Roy. Astron. Soc.*, 396(3):1815–1824, July 2009. doi: 10.1111/j.1365-2966.2009.14885.x.

Furlanello, T., Lipton, Z. C., Tschannen, M., Itti, L., and Anandkumar, A. Born again neural networks. In *International Conference on Machine Learning*, 2018. URL https://api.semanticscholar.org/CorpusID:4110009.

Furniss, A., Sutter, P. M., Primack, J. R., and Domínguez, A. A correlation between hard gamma-ray sources and cosmic voids along the line of sight. *Mon. Not. Roy. Astron. Soc.*, 446(3):2267–2273, January 2015. doi: 10.1093/mnras/stu2196.

Ghafour, P., Tavasoli, S., and Shojaei, M. R. VEGA: Voids idEntification using Genetic Algorithm. *J. Cosmol. Astropart. Phys.*, 2025(12):001, December 2025. doi: 10.1088/1475-7516/2025/12/001.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural message passing for quantum chemistry. 2017. URL https://arxiv.org/abs/1704.01212.

Gori, M., Monfardini, G., and Scarselli, F. A new model for learning in graph domains. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, volume 2, pp. 729–734 vol. 2, 2005. doi: 10.1109/IJCNN.2005.1555942.

Grady, L. Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1768–1783, 2006. doi: 10.1109/TPAMI.2006.233.

Granett, B. R., Neyrinck, M. C., and Szapudi, I. An Imprint of Superstructures on the Microwave Background due to the Integrated Sachs-Wolfe Effect. *Astrophys. J. Lett.*, 683(2):L99, August 2008. doi: 10.1086/591670.

Hamilton, W. L., Ying, R., and Leskovec, J. Inductive representation learning on large graphs. 2018. URL https://arxiv.org/abs/1706.02216.

Horowitz, B. and Bayer, A. E. jFoF: GPU Cluster Finding with Gradient Propagation. *arXiv e-prints*, art. arXiv:2510.26851, October 2025. doi: 10.48550/arXiv.2510.26851.

Horowitz, B., Cuesta-Lazaro, C., and Yehia, O. BaryonBridge: Stochastic Interpolant Model for Fast Hydrodynamical Simulations. *arXiv e-prints*, art. arXiv:2510.19224, October 2025. doi: 10.48550/arXiv.2510.19224.

Hoyle, F. and Vogeley, M. S. Voids in the Point Source Catalogue Survey and the Updated Zwicky Catalog. *Astrophys. J.*, 566(2):641–651, February 2002. doi: 10.1086/338340.

Huchra, J. P. and Geller, M. J. Groups of Galaxies. I. Nearby groups. *Astrophys. J.*, 257:423–437, June 1982. doi: 10.1086/160000.

Hwang, S. Y., Sabiu, C. G., Park, I., and Hong, S. E. The universe is worth 64^3 pixels: convolution neural network and vision transformers for cosmology. *J. Cosmol. Astropart. Phys.*, 2023(11):075, November 2023. doi: 10.1088/1475-7516/2023/11/075.

Kakadia, J., Agrawal, S., Zhong, K., and Jain, B. Vision transformers for cosmological fields: Application to weak lensing mass maps. 2025. URL https://arxiv.org/abs/2512.07125.

Kannan, S., Qiu, T., Cuesta-Lazaro, C., and Jeong, H. CosmoFlow: Scale-Aware Representation Learning for Cosmology with Flow Matching. *arXiv e-prints*, art. arXiv:2507.11842, July 2025. doi: 10.48550/arXiv.2507.11842.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. 2017. URL https://arxiv.org/abs/1412.6980.

Kololgi, D., Naidoo, K., Saintonge, A., and Lahav, O. Learning the Cosmic Web: Graph-based Classification of Simulated Galaxies by their Dark Matter Environments. *arXiv e-prints*, art. arXiv:2512.05909, December 2025. doi: 10.48550/arXiv.2512.05909.

Kuhn, H. W. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955. doi: https://doi.org/10.1002/nav.3800020109. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/nav.3800020109.

Kumagai, S., Vogeley, M. S., Aragon-Calvo, M. A., Douglass, K. A., BenZvi, S., and Neyrinck, M. DeepVoid: A Deep Learning Void Detector. *arXiv e-prints*, art. arXiv:2504.21134, April 2025. doi: 10.48550/arXiv.2504.21134.

Lavaux, G. and Wandelt, B. D. Precision cosmology with voids: definition, methods, dynamics. *Mon. Not. Roy. Astron. Soc.*, 403(3):1392–1408, April 2010. doi: 10.1111/j.1365-2966.2010.16197.x.

Leclercq, F., Jasche, J., Sutter, P. M., Hamaus, N., and Wandelt, B. Dark matter voids in the SDSS galaxy survey. *J. Cosmol. Astropart. Phys.*, 2015(3):047–047, March 2015a. doi: 10.1088/1475-7516/2015/03/047.

Leclercq, F., Jasche, J., and Wandelt, B. Bayesian analysis of the dynamic cosmic web in the SDSS galaxy survey. *J. Cosmol. Astropart. Phys.*, 2015(6):015–015, June 2015b. doi: 10.1088/1475-7516/2015/06/015.

Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Liu, X., Gong, C., and Liu, Q. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. *arXiv e-prints*, art. arXiv:2209.03003, September 2022. doi: 10.48550/arXiv.2209.03003.

Liu, Y. et al. 2026. in prep.

Malandrino, R., Lavaux, G., Wandelt, B. D., McAlpine, S., and Jasche, J. A Bayesian catalog of 100 high-significance voids in the Local Universe. *Astron. Astrophys.*, 705:A160, January 2026. doi: 10.1051/0004-6361/202556345.

Melchior, P., Sutter, P. M., Sheldon, E. S., Krause, E., and Wandelt, B. D. First measurement of gravitational lensing by cosmic voids in SDSS. *Mon. Not. Roy. Astron. Soc.*, 440(4):2922–2927, June 2014. doi: 10.1093/mnras/stu456.

Monllor-Berbegal, Ó., Vallés-Pérez, D., Planelles, S., and Quilis, V. AVISM: Algorithm for Void Identification in coSMology. *Astron. Astrophys.*, 704:A24, November 2025. doi: 10.1051/0004-6361/202554513.

Moresco, M. et al. Unveiling the Universe with emerging cosmological probes. *Living Reviews in Relativity*, 25(1): 6, December 2022. doi: 10.1007/s41114-022-00040-z.

Mousavi-Hosseini, A., Sanford, C., Wu, D., and Erdogdu, M. A. When do transformers outperform feed-forward and recurrent networks? a statistical perspective. In *High-dimensional Learning Dynamics 2025*, 2025. URL https://openreview.net/forum?id=pmVLfyqk69.

Nadathur, S., Carter, P. M., Percival, W. J., Winther, H. A., and Bautista, J. E. REVOLVER: REal-space VOid Locations from suVEy Reconstruction. Astrophysics Source Code Library, record ascl:1907.023, July 2019.

Navarro, J. F., Frenk, C. S., and White, S. D. M. A Universal Density Profile from Hierarchical Clustering. *Astrophys. J.*, 490(2):493–508, December 1997. doi: 10.1086/304888.

Neyrinck, M. C. ZOBOV: a parameter-free void-finding algorithm. *Mon. Not. Roy. Astron. Soc.*, 386(4):2101–2109, June 2008. doi: 10.1111/j.1365-2966.2008.13180.x.

Padilla, N. D., Ceccarelli, L., and Lambas, D. G. Spatial and dynamical properties of voids in a Λ cold dark matter universe. *Mon. Not. Roy. Astron. Soc.*, 363(3):977–990, November 2005. doi: 10.1111/j.1365-2966.2005.09500.x.

Paillas, E., Cautun, M., Li, B., Cai, Y.-C., Padilla, N., Armijo, J., and Bose, S. The Santiago-Harvard-Edinburgh-Durham void comparison II: unveiling the Vainshtein screening using weak lensing. *Mon. Not. Roy. Astron. Soc.*, 484(1):1149–1165, March 2019. doi: 10.1093/mnras/stz022.

Pandey, S., Lanusse, F., Modi, C., and Wandelt, B. D. Teaching dark matter simulations to speak the halo language. *arXiv e-prints*, art. arXiv:2409.11401, September 2024. doi: 10.48550/arXiv.2409.11401.

Pandey, S., Lovell, C. C., Modi, C., and Wandelt, B. D. Galactification: painting galaxies onto dark matter only simulations using a transformer-based model. *arXiv e-prints*, art. arXiv:2511.08438, November 2025a. doi: 10.48550/arXiv.2511.08438.

Pandey, S., Modi, C., Wandelt, B. D., Bartlett, D. J., Bayer, A. E., Bryan, G. L., Ho, M., Lavaux, G., Makinen, T. L., and Villaescusa-Navarro, F. Creating halos with autoregressive multistage networks. *Phys. Rev. D*, 112(10): 103503, November 2025b. doi: 10.1103/vlm2-tm6k.

Paszke, A. et al. Pytorch: An imperative style, high-performance deep learning library. 2019. URL https://arxiv.org/abs/1912.01703.

Paz, D. J., Correa, C. M., Gualpa, S. R., Ruiz, A. N., Bederián, C. S., Graña, R. D., and Padilla, N. D. Guess the cheese flavour by the size of its holes: a cosmological test using the abundance of popcorn voids. *Monthly Notices of the Royal Astronomical Society*, 522(2):2553–2569, 04 2023. ISSN 0035-8711. doi: 10.1093/mnras/stad1146. URL https://doi.org/10.1093/mnras/stad1146.

Perez, E., Strub, F., de Vries, H., Dumoulin, V., and Courville, A. Film: Visual reasoning with a general conditioning layer. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), Apr. 2018. doi: 10.1609/aaai.v32i1.11671. URL https://ojs.aaai.org/index.php/AAAI/article/view/11671.

Pisani, A. et al. Cosmic voids: a novel probe to shed light on our Universe. *Bull. Am. Astron. Soc.*, 51(3):40, May 2019. doi: 10.48550/arXiv.1903.05161.

Pomarède, D., Hoffman, Y., Courtois, H. M., and Tully, R. B. The Cosmic V-Web. *Astrophys. J.*, 845(1):55, August 2017. doi: 10.3847/1538-4357/aa7f78.

Pooladian, A.-A., Ben-Hamu, H., Domingo-Enrich, C., Amos, B., Lipman, Y., and Chen, R. T. Q. Multisample flow matching: straightening flows with minibatch couplings. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

Robotham, A. S. G. and Howlett, C. A Short Research Note on Calculating Exact Distribution Functions and Random Sampling for the 3D NFW Profile. *Research Notes of the American Astronomical Society*, 2(2):55, June 2018. doi: 10.3847/2515-5172/aacc70.

Ronconi, T. and Marulli, F. Cosmological exploitation of cosmic void statistics. New numerical tools in the CosmoBolognaLib to extract cosmological constraints from the void size function. *Astron. Astrophys.*, 607:A24, October 2017. doi: 10.1051/0004-6361/201730852.

Ruiz, A. N., Paz, D. J., Lares, M., Luparello, H. E., Ceccarelli, L., and Lambas, D. G. Clues on void evolution - III. Structure and dynamics in void shells. *Mon. Not. Roy. Astron. Soc.*, 448(2):1471–1482, April 2015. doi: 10.1093/mnras/stv019.

Ruiz, A. N., Alfaro, I. G., and Garcia Lambas, D. Structure and dynamics in low-density regions: galaxy-galaxy correlations inside cosmic voids. *Mon. Not. Roy. Astron. Soc.*, 483(3):4070–4079, March 2019. doi: 10.1093/mnras/sty3443.

Sartori, S., Contarini, S., Sarpa, E., Degni, G., Marulli, F., Escoffier, S., and Moscardini, L. The back-in-time void finder: dynamical identification of cosmic voids through optimal transport reconstruction. 2026. URL https://arxiv.org/abs/2601.15378.

Sartori, S. et al. The imprint of cosmic voids from the DESI Legacy Survey DR9 Luminous Red Galaxies in the Planck 2018 lensing map through spectroscopically calibrated mocks. *Astron. Astrophys.*, 700:A17, August 2025. doi: 10.1051/0004-6361/202453562.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009. doi: 10.1109/TNN.2008.2005605.

Schmidt, J. D., Ryden, B. S., and Melott, A. L. The Size and Shape of Voids in Three-dimensional Galaxy Surveys. *Astrophys. J.*, 546(2):609–619, January 2001. doi: 10.1086/318313.

Schuster, N., Hamaus, N., Dolag, K., and Weller, J. Why cosmic voids matter: nonlinear structure & linear dynamics. *J. Cosmol. Astropart. Phys.*, 2023(5):031, May 2023. doi: 10.1088/1475-7516/2023/05/031.

Schuster, N., Hamaus, N., Dolag, K., and Weller, J. Why cosmic voids matter: mitigation of baryonic physics. *J. Cosmol. Astropart. Phys.*, 2024(8):065, August 2024. doi: 10.1088/1475-7516/2024/08/065.

Schuster, N., Hamaus, N., Pisani, A., Dolag, K., and Weller, J. Why Cosmic Voids Matter: Pristine Evolution. *arXiv e-prints*, art. arXiv:2509.07092, September 2025. doi: 10.48550/arXiv.2509.07092.

Sousbie, T. DisPerSE: Discrete Persistent Structures Extractor. Astrophysics Source Code Library, record ascl:1302.015, February 2013.

Sutter, P. M., Lavaux, G., Hamaus, N., Wandelt, B. D., Weinberg, D. H., and Warren, M. S. Sparse sampling, galaxy bias, and voids. *Mon. Not. Roy. Astron. Soc.*, 442(1):462–471, July 2014. doi: 10.1093/mnras/stu893.

Sutter, P. M., Lavaux, G., Hamaus, N., Pisani, A., Wandelt, B. D., Warren, M., Villaescusa-Navarro, F., Zivick, P., Mao, Q., and Thompson, B. B. VIDE: The Void IDentification and Examination toolkit. *Astronomy and Computing*, 9:1–9, March 2015. doi: 10.1016/j.ascom.2014.10.002.

Tong, A., Fatras, K., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Wolf, G., and Bengio, Y. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=CD9Snc73AW.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. Graph attention networks. 2018. URL https://arxiv.org/abs/1710.10903.

Verza, G., Degni, G., Pisani, A., Hamaus, N., Massara, E., Benson, A., Escoffier, S., Wang, Y., Zhai, Z., and Doré, O. Cosmology with Voids from the Nancy Grace Roman Space Telescope. *Astrophys. J.*, 993(2):227, November 2025. doi: 10.3847/1538-4357/ae07d9.

Villaescusa-Navarro, F. et al. The Quijote Simulations. *Astrophys. J. Suppl.*, 250(1):2, September 2020. doi: 10.3847/1538-4365/ab9d82.

Wechsler, R. H. and Tinker, J. L. The Connection Between Galaxies and Their Dark Matter Halos. *Annu. Rev. Astron. Astrophys.*, 56:435–487, September 2018. doi: 10.1146/annurev-astro-081817-051756.

Xie, Q., Luong, M.-T., Hovy, E., and Le, Q. V. Self-training with noisy student improves imagenet classification. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, 2020. doi: 10.1109/CVPR42600.2020.01070.

Zhao, C., Tao, C., Liang, Y., Kitaura, F.-S., and Chuang, C.-H. DIVE in the cosmic web: voids with Delaunay triangulation from discrete matter tracer distributions. *Mon. Not. Roy. Astron. Soc.*, 459(3):2670–2680, July 2016. doi: 10.1093/mnras/stw660.

Zheng, Z., Berlind, A. A., Weinberg, D. H., Benson, A. J., Baugh, C. M., Cole, S., Davé, R., Frenk, C. S., Katz, N., and Lacey, C. G. Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies. *Astrophys. J.*, 633(2):791–809, November 2005. doi: 10.1086/466510.

A. You *can* have an appendix here.

B. Halo Occupation Distribution

The central galaxy is placed at the halo center with probability

$$\overline{N}_{\text{cen}} = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\log M - \log M_{\text{min}}}{\sigma_{\log M}}\right)\right] , \tag{10}$$

and the number of satellites is drawn from a Poisson distribution with mean

$$\overline{N}_{\text{sat}} = \overline{N}_{\text{cen}} \left(\frac{M - M_0}{M_1}\right)^{\alpha} . \tag{11}$$

The satellites are distributed isotropically according to an NFW profile (Navarro et al., 1997) and the concentration model of Duffy et al. (2008), using the analytic solution for the inverse NFW CDF from Robotham & Howlett (2018).

C. Feature engineering

The edge features are taken as

$$e^{GP}_{\alpha\beta} = \{\Delta\vec{r}_{\alpha\beta}, |\Delta\vec{r}_{\alpha\beta}|\} , \tag{12}$$

and analogous for $e^{PP}_{\alpha\alpha'}$.

For the node features we aim to approximately respect translational symmetry by projecting the coordinates on a Fourier basis. Specifically, we compute (in the natural interpretation of the notation)

$$x^P_{\alpha,r} = \{\sin, \cos\}(L^{-1}\pi\nu\vec{r}^P_{\alpha}) , \nu = 2^{0\ldots m-1} , \tag{13}$$

where $L = 1\,h^{-1}$Gpc is the box size and $m = 14$. The transformation

$$x^P_{\alpha,r} \leftarrow \text{erf}^{-1}(2\pi^{-1} \arcsin x^P_{\alpha,r}) \tag{14}$$

Gaussianizes these Fourier features. Furthermore, we compute a number of summary statistics of the local particle distribution. This is particularly important for the galaxy nodes as these do not receive updates in message passing. Using the set

$$d_\alpha = \{\Delta\vec{r}_{\alpha\beta}\}_{\beta\in\text{knn}(\alpha)} \tag{15}$$

and the corresponding sets s_α of vector lengths and u_α of unit vectors, we compute the density proxies

$$x^P_{\alpha,\delta} = \{\text{max}, \text{mean}, \text{std}\}(s_\alpha) , \tag{16}$$

the gradient proxies

$$x^P_{\alpha,\nabla} = \sum u_\alpha , \tag{17}$$

and from the mass matrix

$$\sum d_\alpha \otimes d_\alpha \tag{18}$$

the eigenvalues $x^P_{\alpha,\lambda}$ and major axis $x^P_{\alpha,v}$. The full feature vector

$$x^P_\alpha = \{x^P_{\alpha,r}, x^P_{\alpha,\delta}, x^P_{\alpha,\nabla}, x^P_{\alpha,\lambda}, x^P_{\alpha,v}\} \tag{19}$$

has dimension 96. The construction of galaxy node features x^G_β is analogous. We whiten all node and edge features to zero mean and unit variance.

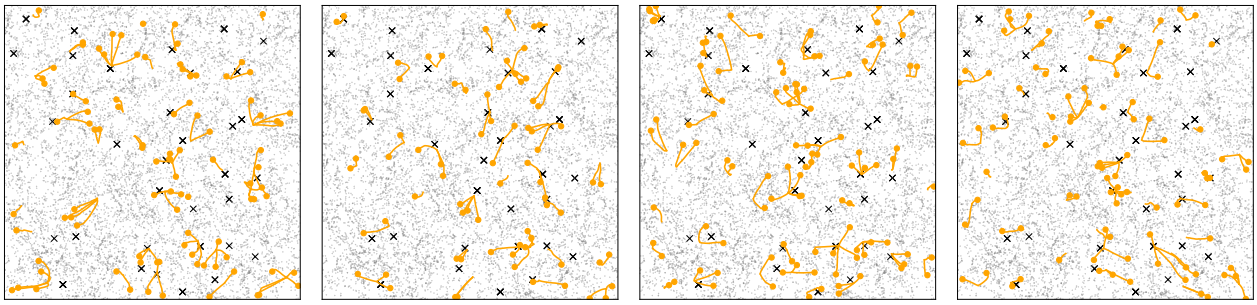

Figure 10. Variation of trajectories with test particle initial conditions, compare to Fig. 3. Each panel is run with identical galaxies but varying random seed for the test particle initial conditions. We observe substantial variation in the trajectory endpoints.

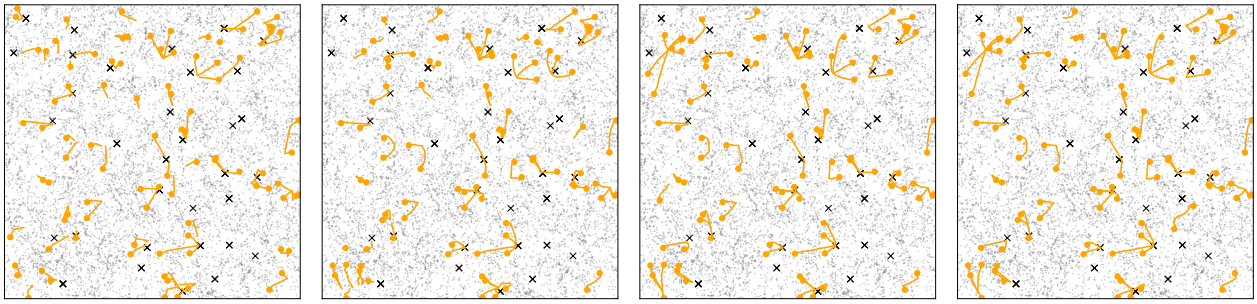

Figure 11. Variation of trajectories with integration accuracy, compare to Fig. 3. From left to right, we use 2, 8, 32 (fiducial choice), and 128 time steps in the integration of the flow ODE Eq. (3) while keeping the random seed fixed. The convergence with number of integration steps is good.

D. Example test particle trajectories

In order to develop further intuition for the variation of output void catalogs with different initial conditions, and thus the level of stochasticity learned by our network, in Fig. 10 we show analogs of Fig. 3 when the test particle initial conditions are varied.

In Fig. 11 we check for convergence of test particle trajectories with respect to the number of integration steps. The fiducial choice of 32 steps adopted in the main text is seen to be extremely well converged. This is a consequence of the network's being well-trained and thus predicting low-curvature trajectories.