Galaxies on graph neural networks: towards robust synthetic galaxy catalogs with deep generative models.

Yesukhei Jagvaral [1 2] **François Lanusse** [3] **Sukhdeep Singh** [1 2] **Rachel Mandelbaum** [1 2] **Siamak Ravanbakhsh** [4] **Duncan Campbell** [1 5]

Abstract

The future astronomical imaging surveys are set to provide precise constraints on cosmological parameters, such as dark energy. However, production of synthetic data for these surveys, to test and validate analysis methods, suffers from a very high computational cost. In particular, generating mock galaxy catalogs at sufficiently large volume and high resolution will soon become computationally unreachable. In this paper, we address this problem with a Deep Generative Model to create robust mock galaxy catalogs that may be used to test and develop the analysis pipelines of future weak lensing surveys. We build our model on a custom built Graph Convolutional Networks, by placing each galaxy on a graph node and then connecting the graphs within each gravitationally bound system. We train our model on a cosmological simulation with realistic galaxy populations to capture the 2D and 3D orientations of galaxies. The samples from the model exhibit comparable statistical properties to those in the simulations. To the best of our knowledge, this is the first instance of a generative model on graphs in an astrophysical/cosmological context

[1]McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA [2]NSF AI Planning Institute for Data-Driven Discovery in Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA [3]AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, F-91191 Gif-sur-Yvette, France [4]Mila, Quebec AI Institute [5]Epistemix Inc., Pittsburgh. Correspondence to: Yesukhei Jagvaral <yjagvara@andrew.cmu.edu>.

ICML 2022 Workshop on Machine Learning for Astrophysics, Baltimore, Maryland, USA, 2022.

1. Introduction

Upcoming astronomical imaging surveys such as the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST)[1], Roman Space Telescope[2] High Latitude Survey (HLS) and Euclid[3] will aim to answer fundamental questions about the nature of dark matter and dark energy, by precisely measuring the distribution and properties of billions of galaxies.

The analysis of these surveys will require having an access to large scale cosmological simulations for a variety of applications, ranging from validating analysis pipelines (DeRose et al., 2019; 2021) to constraining cosmology through Simulation-Based Inference (SBI; Jeffrey et al., 2021). However, as the volume and data quality of future surveys increases, cosmological simulation must cover increasingly large volumes at high resolution (Vogelsberger et al., 2020). Full hydrodynamical simulations, which can resolve the formation and evolution of individual galaxies, are extremely expensive and cannot scale to such volumes. This motivates the need for emulation methods capable of generating realistic mock galaxy catalogs without relying on a full simulation. One traditional solution to this problem has been to (semi-)analytically paint galaxies on N-body (gravity only) simulations. However, the assumptions behind these (semi-)analytical models may not be robust and need validation from non-parametric models (Somerville & Davé, 2015). While machine learning would be an appealing solution to this problem, One of the main difficulties in building ML-based non-parametric models for such simulations is the fact that the data to emulate is catalog-based (i.e., a catalog of galaxy positions and properties in the simulation volume) and that each galaxy cannot be treated independently if important correlations between galaxies are to be preserved.

In this work, we propose to address this emulation problem with a conditional deep generative model, capable of model-

[1]https://www.lsst.org/
[2]https://roman.gsfc.nasa.gov/
[3]https://www.euclid-ec.org/

ing relevant galaxy properties and their inter-dependencies, conditioned on the underlying large-scale structure scaffolding. Our model combines a customized graph convolutional network architecture, to model the correlations between galaxies, with a Wasserstein Generative Adversarial Network to build a probabilistic model of galaxy properties. To the best of our knowledge, this work is the first instance of a deep generative model on graphs introduced in astrophysics.

We apply our model to the particularly challenging problem of modeling the intrinsic alignments of galaxies. The model is able to learn and predict both (a) scalar features such as galaxy shapes, and more importantly, (b) correlated vector orientations in 3D and in 2D to a good quantitative agreement.

2. Related Work

In the literature there is substantial work on galaxy property emulators. In some cases (Agarwal et al., 2018; Modi et al., 2018; Zhang et al., 2019) the approach has been to 'paint' galaxy properties onto N-body (dark matter – gravity only) simulations. However, these methods typically do not model correlations between galaxies and only predict scalar quantities, as opposed to our model which predicts both vector and scalar quantities.

While graph neural networks have been proposed in the context of cosmological simulations in previous work, it is the first time that they are used to build generative models for galaxy properties. Cranmer et al. (2020) trained a graph neural network on a cosmological simulation and then extracted symbolic equations pertaining to physical laws. Villanueva-Domingo et al. (2021), on the other hand, used graph neural networks to infer halo masses.

3. Directional Graph Convolutional Networks

A key feature of our problem is that the neural architecture needs to model direction- and distance-dependent correlations between galaxies. Instead of relying on a generic message-passing approach to build a graph neural network, we use our physical insight to build an architecture with explicit dependence on relative distance and orientation between graph nodes, as described below. In this work, we are considering undirected and connected graphs: $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, where $\mathcal{V}$ is the set of graphs vertices, with $|\mathcal{V}| = n$ the number of vertices, $\mathcal{E}$ is the set of graph edges and $\mathbf{W} \in \mathbb{R}^{n \times n}$ is the weighted adjacency matrix. We adopt a first order approximation to parameterize graph convolutions (Kipf & Welling, 2016), and define one Graph Convolutional Network layer with an activation y_i for a node i as:

$$\forall i \in \mathcal{V}, \;\; y_i = \mathbf{b} + \mathbf{W}_0 h_i + \sum_{j \in \mathcal{N}_i} w_{i,j} \mathbf{W}_1 h_j \tag{1}$$

where $\mathbf{b}$ represents a vector of bias terms; $\mathcal{N}_i$ denotes the set of immediate neighbors[4] of vertex i; $\mathbf{W}_0$ are the weights that apply a linear transform to the activation vector h_i of node i (i.e., self connection); $w_{i,j}$ are linear transforms on the activation vectors h_j of the nodes j in the neighborhood of i; and $\mathbf{W}_1$ are the set of weights that apply to the immediate neighbors.

Following an approach proposed in Verma et al. (2017), we implement direction-dependent graph convolution layer as

$$y_i = \mathbf{b} + \mathbf{W}_0 h_i + \sum_{m=1}^{M} \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} q_m(\mathbf{r}_i, \mathbf{r}_j) \mathbf{W}_m h_j . \tag{2}$$

Here $|\mathcal{N}_i|$ denotes the cardinality of the set $\mathcal{N}_i$, M is the number of directions, and $\mathbf{r}_i$ are the 3D Cartesian coordinates of the node. The $q_m(\mathbf{r}_i, \mathbf{r}_j)$ are normalized so that $\sum_{m=1}^{M} q_m(\mathbf{r}_i, \mathbf{r}_j) = 1$ and are defined as:

$$q_m(\mathbf{r}_i, \mathbf{r}_j) \propto \exp\left(-\mathbf{d}_m^t \cdot (\mathbf{r}_j - \mathbf{r}_i)\right) \, g_\lambda(\| \mathbf{r}_i - \mathbf{r}_j \|_2^2), \tag{3}$$

where the $\{\mathbf{d}_m\}_{m \in [1,M]}$ are a set of directions we want to make the kernel sensitive to, and g_λ is a parametric function of the distance between two vertices. This can be seen as a hard-coded direction-dependent attention mechanism allowing the model to gain directional awareness by design. We further chose an exponential parametrization of the form $g_\lambda(r) = \exp\left(-r^2/2\lambda^2\right)$ for the distance-dependence, where λ is fit automatically during training. Note that more generic functions could be used, but this empirical parametrization was found to work well for our problem.

4. Generative Model with Graphs

Our goal is to learn and sample from a conditional probability density $p(\boldsymbol{y}|\boldsymbol{x})$, where $\boldsymbol{y}$ might be an orientation of a galaxy, and $\boldsymbol{x}$ would be quantities such as the dark matter mass of a galaxy or the tidal field at its location. We model this distribution by employing a conditional Wasserstein generative adversarial network (GAN, Goodfellow et al. 2014). GANs were chosen to model complex joint probability densities of all galaxies in a halo, without needing a parametric form/probability model.

Given a generating function $g_\theta(z, \boldsymbol{x})$ with $z \sim \mathcal{N}(0, \boldsymbol{I})$, we aim to adjust the implicit distribution generated by g_θ to match our target distribution $p(\boldsymbol{y}|\boldsymbol{x})$. This can be done by minimizing the Wasserstein 1-distance $\mathcal{W}$ between these two distributions to find an optimal set of weights $\theta_\star$. By using an approximate Wasserstein distance, we are solving

[4] Immediate neighbors or first neighbors are neighbors that are one hop away from node i.

the following minimax optimization problem:

$$\arg\min_{\theta}\left(\sup_{\phi}\mathbb{E}_{(x,y)}\left[d_{\phi}(\boldsymbol{x},\boldsymbol{y})-\mathbb{E}_{z}\left[d_{\phi}(g_{\theta}(z,\boldsymbol{x}),\boldsymbol{y})\right]\right]\right) \tag{4}$$

Additionally, we must keep the Lipschitz constant bounded, to ensure that d_ϕ indeed parameterizes a Wasserstein distance. In Arjovsky et al. (2017), the authors have clipped the weights of the model to ensure the Lipschitz condition. Later Gulrajani et al. (2017) showed that the gradient constraint performs better – thus in this work we adopt a gradient penalty.

5. Application: Emulating Galaxy Intrinsic Alignments in Illustris-TNG simulations

Images of distant galaxies come to us with distortions, excluding camera and atmospheric effects. These distortions are caused by a phenomena known as weak gravitational lensing, where light traveling from the galaxy gets deflected due to the presence of a massive objects (like a galaxy cluster) on the light's pathway. Weak lensing is measured using statistical ensembles of galaxies and their coherent shape distortions, which are caused by the matter distribution in the Universe coherently distorting space-time. However, weak lensing measurements suffer from a number of systematics, one of which is intrinsic alignments - where galaxies are not oriented randomly in the sky in the absence of weak lensing effects, but rather tend to point towards dense regions, including those hosting other galaxies. This effect contaminates our desired weak lensing signal and can bias our measurement of dark energy. Realistic modeling of these alignments in mock galaxy catalogs is therefore paramount to validate the robustness of analysis pipelines.

5.1. Simulated data

In this work we are using the hydrodynamical TNG100-1 run from the IllustrisTNG simulation suite (for more information, please refer to Nelson et al., 2018; Pillepich et al., 2018; Springel et al., 2018; Naiman et al., 2018; Marinacci et al., 2018; Nelson et al., 2019). We employ a minimum stellar mass threshold of $\log_{10}(M_*/M_\odot) = 9$ for all galaxies, using their stellar mass from the SUBFIND catalog.

5.2. Graph construction

To construct the graph for the cosmic web (i.e., for the subhalos and the galaxies), we first grouped all subhalos and galaxies based on their parent halo. Given a galaxy catalog, an undirected graph based on the 3D positions within the parent halo is built by placing each galaxy on a *graph node*. Then, each node has a list of features such as mass and tidal field. To build the graph connection for a given group, the nearest neighbors within a specified radius for a given node are connected via the *undirected edges* with *signals* on the graphs representing the alignments. A snapshot of the simulation represented as *graphs* is shown in Fig. 1, where each node represents a galaxy and the connections between the nodes are represented by grey lines.

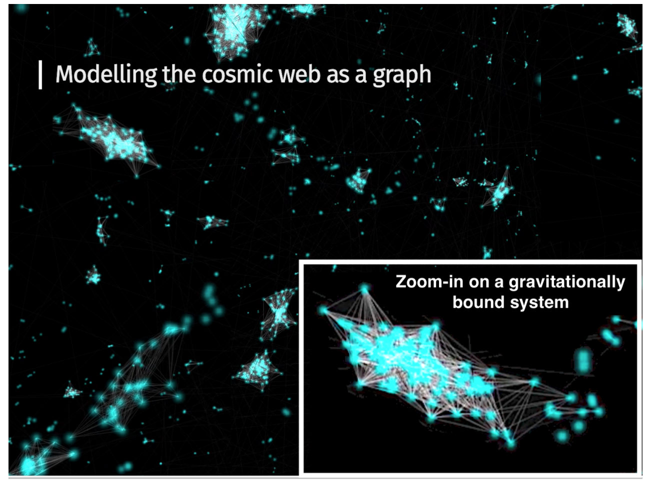

Figure 1. The cosmological simulation box modeled as a graph. Here every node represents a galaxy and the connections are made using the r-NN algorithm within each gravitationally bound system. Graph animation adapted from Coutinho et al. (2016).

5.3. Model Architecture

In Fig. 2 we outline the architecture of our model. We have list of features (orange box) that are relevant for capturing the dependence of intrinsic alignments within a halo (dashed red box), and the tidal fields that are relevant for capturing the dependence of IA for galaxies on matter beyond their halo (dashed purple). These inputs are fed into the GAN-Generator (crimson box), which tries to learn the statistical distribution of the desired output labels (yellow box). At the end the input and the output from the GAN-Generator are fed into the GAN-Critic (blue box) to determine the performance of the GAN-Generator. In our model, the Generator has 5 layers each with {128,128,16,2,2} neurons, while the Critic has 4 layers each with {128,128,64,32} neurons followed by a mean-pooling layer and a single output neuron.

5.4. Training

We train the model using the Adam optimizer (Kingma & Ba, 2014) with a learning rate of 10^{-3} and exponential decay rates of $\beta_1 = 0$ and $\beta_2 = 0.95$. During the adversarial training we train the Generator for 5 steps and the Critic for 1 step with a batch size of 64 (one batch is set of graphs) and a leaky ReLU activation function. As is common with

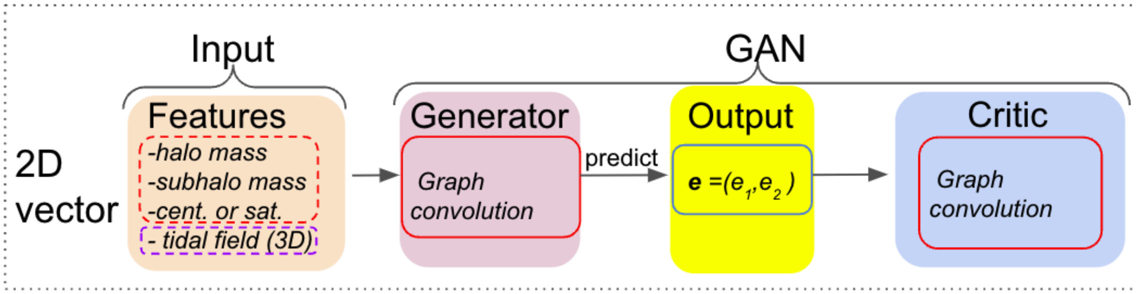

Figure 2. Architecture of the graph convolution GAN model used. Here, the input features are typical of medium to high resolution N-body (gravity only) simulations. The $\mathbf{e} = (e_1, e_2)$ is the 2D quantity that parametrizes the galaxy orientation in the sky.

GANs, our GAN models do not converge; we arbitrarily stop the training once it reached a reasonable result. Our code is available at [].

5.5. Results

Throughout the section we refer to the sample generated from the Graph-Convolutional Network-based Generative Adversarial Networks as the *GAN* sample, and the sample from the TNG100 simulation as the *TNG* sample.

For our key result, we examine w_{g+}, the density-shape correlation function computed using the ellipticities (can be thought of as the 2D orientation and flattening of a galaxy when modeled as an ellipse), as shown in Fig. 3. The projected density-shape correlation function w_{g+} captures the correlation between overdensity and projected intrinsic ellipticity, as is commonly used in observational studies. Positive values for w_{g+} indicate that galaxies exhibit a coherent alignment towards the locations of other nearby galaxies. We split our sample roughly 50/50 into training and testing samples, while preserving group membership of subhalos and galaxies. The projected 2D correlation function, w_{g+}, from the GAN agrees quantitatively with the measured one from TNG simulation. Here, the errorbars were derived from an analytic estimate of the covariance matrix, which includes Gaussian terms for noise and cosmic variance (for more details see Singh et al. 2017; Samuroff et al. 2020). Additionally, our model is also able to predict 3D orientations and scalar quantities to a similar level of quantitative agreement.

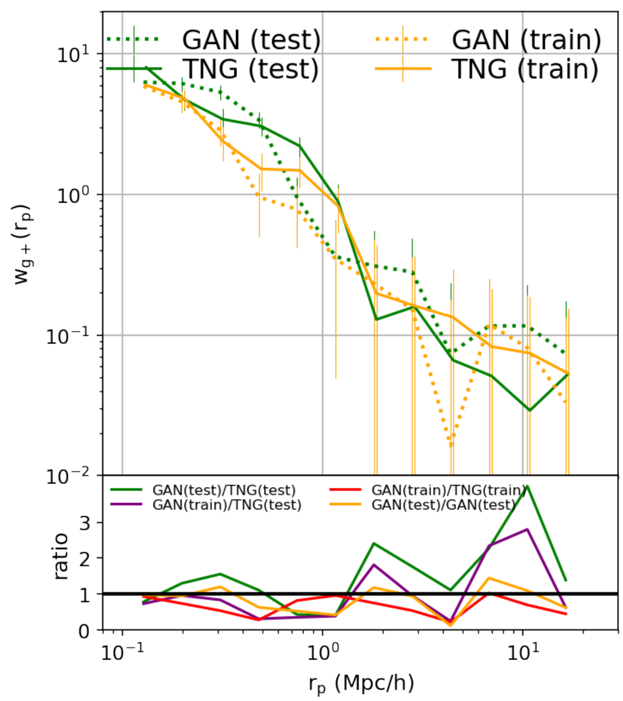

Figure 3. Projected two-point correlation functions w_{g+} of galaxy positions and the projected 2D ellipticities of all galaxies, split into roughly equal-sized training and testing samples while preserving group membership. The top panel shows w_{g+} as a function of projected galaxy separation r_p measured using data from the TNG simulation in yellow and the data generated by the GAN in dotted green, while the bottom panel shows the ratios among the curves as indicated by the label. All four curves are in good quantitative agreement, suggesting that the GAN is not significantly overfitting.

6. Conclusions

In this abstract, we have presented a novel deep generative model for scalar and vector galaxy properties. Using the TNG100 hydrodynamical simulation from the IllustrisTNG simulation suite, we have trained the model to accurately predict galaxy orientations. For a simulation box of 75 Mpc/h with 20k galaxies, the training takes about 3–4 days on a modern GPU; applying the model to a dataset of equal size is very fast (less than a minute).

Overall, the Graph Convolution based Generative Adversarial network learns and generates scalar and vector quantities that have statistical properties (distributions and alignment correlations) that agree well with those of the original simulation. Learning galaxy orientations is part of a more general problem called Galaxy-Halo connection. The problem can be stated as follows: given some properties of a dark matter halo can we predict what type of galaxy it hosts, or vice versa? Our results represent a concrete step towards addressing this complex problem with Graph Neural Network-based Deep Generative Models.

Future work includes applying this model on a much higher volume N-body simulation with lower resolution in order to utilize the power of deep generative models for upcoming cosmological surveys. Additionally, incorporating symmetries such as SO(3) or E(3) and making equivariant neural networks for graphs (Horie et al., 2020; Satorras et al., 2021) would be a useful development.

Acknowledgements

We thank Ananth Tenneti, Tiziana DiMatteo, Barnabas Poczos and Rupert Croft for useful discussion that informed the direction of this work. This work was supported in part by the National Science Foundation, NSF AST-1716131 and by a grant from the Simons Foundation (Simons Investigator in Astrophysics, Award ID 620789). SS is supported by a McWilliams postdoctoral fellowship at Carnegie Mellon University. This work is supported by the NSF AI Institute: Physics of the Future, NSF PHY- 2020295.

References

Agarwal, S., Davé, R., and Bassett, B. A. Painting galaxies into dark matter haloes using machine learning. *Monthly Notices of the Royal Astronomical Society*, 478(3):3410–3422, August 2018. doi: 10.1093/mnras/sty1169.

Arjovsky, M., Chintala, S., and Bottou, L. Wasserstein GAN. *arXiv e-prints*, art. arXiv:1701.07875, January 2017.

Coutinho, B. C., Hong, S., Albrecht, K., Dey, A., Barabási, A.-L., Torrey, P., Vogelsberger, M., and Hernquist, L. The Network Behind the Cosmic Web. *arXiv e-prints*, art. arXiv:1604.03236, April 2016.

Cranmer, M., Sanchez-Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D., and Ho, S. Discovering Symbolic Models from Deep Learning with Inductive Biases. *arXiv e-prints*, art. arXiv:2006.11287, June 2020.

DeRose, J., Wechsler, R. H., Becker, M. R., Busha, M. T., Rykoff, E. S., MacCrann, N., Erickson, B., Evrard, A. E., Kravtsov, A., Gruen, D., Allam, S., Avila, S., Bridle, S., Brooks, D., Buckley-Geer, E., Carnero Rosell, A., Carrasco Kind, M., Carretero, J., Castander, F. J., Cawthon, R., Crocce, M., da Costa, L. N., Davis, C., De Vicente, J., Dietrich, J. P., Doel, P., Drlica-Wagner, A., Fosalba, P., Frieman, J., Garcia-Bellido, J., Gutierrez, G., Hartley, W. G., Hollowood, D. L., Hoyle, B., James, D. J., Krause, E., Kuehn, K., Kuropatkin, N., Lima, M., Maia, M. A. G., Menanteau, F., Miller, C. J., Miquel, R., Ogando, R. L. C., Plazas Malagón, A., Romer, A. K., Sanchez, E., Schindler, R., Serrano, S., Sevilla-Noarbe, I., Smith, M., Suchyta, E., Swanson, M. E. C., Tarle, G., and Vikram, V. The Buzzard Flock: Dark Energy Survey Synthetic Sky Catalogs. *arXiv e-prints*, art. arXiv:1901.02401, January 2019.

DeRose, J., Wechsler, R. H., Becker, M. R., Rykoff, E. S., Pandey, S., MacCrann, N., Amon, A., Myles, J., Krause, E., Gruen, D., Jain, B., Troxel, M. A., Prat, J., Alarcon, A., Sánchez, C., Blazek, J., Crocce, M., Giannini, G., Gatti, M., Bernstein, G. M., Zuntz, J., Dodelson, S., Fang, X., Friedrich, O., Secco, L. F., Elvin-Poole, J., Everett, S., Choi, A., Harrison, I., Cordero, J., Rodriguez-Monroy, M., McCullough, J., Cawthon, R., Chen, A., Alves, O., Camacho, H., Campos, A., Diehl, H. T., Drlica-Wagner, A., Eifler, T. F., Fosalba, P., Huang, H., Porredon, A., Raveri, M., Rosenfeld, R., Ross, A. J., Sanchez, J., Sheldon, E., Yanny, B., Yin, B., Aguena, M., Allam, S., Andrade-Oliveira, F., Annis, J., Avila, S., Bacon, D., Bechtol, K., Bhargava, S., Brooks, D., Buckley-Geer, E., Burke, D. L., Carnero Rosell, A., Carrasco Kind, M., Chang, C., Costanzi, M., da Costa, L. N., Pereira, M. E. S., De Vicente, J., Desai, S., Dietrich, J. P., Doel, P., Eckert, K., Evrard, A. E., Ferrero, I., Ferté, A., Flaugher, B., Frieman, J., García-Bellido, J., Gaztanaga, E., Giannantonio, T., Gruendl, R. A., Gschwend, J., Gutierrez, G., Hartley, W. G., Hinton, S. R., Hollowood, D. L., Honscheid, K., Huff, E. M., Huterer, D., James, D. J., Kuehn, K., Kuropatkin, N., Lahav, O., Lima, M., Maia, M. A. G., Marshall, J. L., Melchior, P., Menanteau, F., Miquel, R., Mohr, J. J., Morgan, R., Palmese, A., Paz-Chinchón, F., Pieres, A., Plazas Malagón, A. A., Sanchez, E., Scarpine, V., Serrano, S., Sevilla-Noarbe, I., Smith, M., Soares-Santos, M., Suchyta, E., Tarle, G., Thomas, D., To, C., Varga, T. N., and Zhang, Y. Dark Energy Survey Year 3 results: cosmology from combined galaxy clustering and lensing – validation on cosmological simulations. *arXiv e-prints*, art. arXiv:2105.13547, May 2021.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative Adversarial Networks. *arXiv e-prints*, art. arXiv:1406.2661, June 2014.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. Improved Training of Wasserstein GANs. *arXiv e-prints*, art. arXiv:1704.00028, March 2017.

Horie, M., Morita, N., Hishinuma, T., Ihara, Y., and Mitsume, N. Isometric Transformation Invariant and Equivariant Graph Convolutional Networks. *arXiv e-prints*, art. arXiv:2005.06316, May 2020.

Jeffrey, N., Alsing, J., and Lanusse, F. Likelihood-free inference with neural compression of DES SV weak lensing map statistics. *Monthly Notices of the Royal Astronomical Society*, 501(1):954–969, February 2021. doi: 10.1093/mnras/staa3594.

Kingma, D. P. and Ba, J. Adam: A Method for Stochastic Optimization. *arXiv e-prints*, art. arXiv:1412.6980, December 2014.

Kipf, T. N. and Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv e-prints*, art. arXiv:1609.02907, September 2016.

Marinacci, F. et al. First results from the IllustrisTNG simulations: radio haloes and magnetic fields. *Mon. Not. Roy. Astron. Soc.*, 480(4):5113–5139, 2018. doi: 10.1093/mnras/sty2206.

Modi, C., Feng, Y., and Seljak, U. Cosmological reconstruction from galaxy light: neural network based light-matter connection. *Journal of Cosmology and Astroparticle Physics*, 2018(10):028, October 2018. doi: 10.1088/1475-7516/2018/10/028.

Naiman, J. P., Pillepich, A., Springel, V., Ramirez-Ruiz, E., Torrey, P., Vogelsberger, M., Pakmor, R., Nelson, D., Marinacci, F., Hernquist, L., Weinberger, R., and Genel, S. First results from the IllustrisTNG simulations: a tale of two elements - chemical evolution of magnesium and europium. *Monthly Notices of the Royal Astronomical Society*, 477(1):1206–1224, June 2018. doi: 10.1093/mnras/sty618.

Nelson, D., Springel, V., Pillepich, A., Rodriguez-Gomez, V., Torrey, P., Genel, S., Vogelsberger, M., Pakmor, R., Marinacci, F., Weinberger, R., Kelley, L., Lovell, M., Diemer, B., and Hernquist, L. The IllustrisTNG simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1):2, May 2019. doi: 10.1186/s40668-019-0028-x.

Nelson, D. et al. First results from the IllustrisTNG simulations: the galaxy colour bimodality. *Mon. Not. Roy. Astron. Soc.*, 475(1):624–647, 2018. doi: 10.1093/mnras/stx3040.

Pillepich, A., Nelson, D., Hernquist, L., Springel, V., Pakmor, R., Torrey, P., Weinberger, R., Genel, S., Naiman, J. P., Marinacci, F., and Vogelsberger, M. First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies. *Monthly Notices of the Royal Astronomical Society*, 475(1):648–675, March 2018. doi: 10.1093/mnras/stx3112.

Samuroff, S., Mandelbaum, R., and Blazek, J. Advances in Constraining Intrinsic Alignment Models with Hydrodynamic Simulations. *arXiv e-prints*, art. arXiv:2009.10735, September 2020.

Satorras, V. G., Hoogeboom, E., and Welling, M. E(n) equivariant graph neural networks. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9323–9332. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/satorras21a.html.

Singh, S., Mandelbaum, R., Seljak, U., Slosar, A., and Vazquez Gonzalez, J. Galaxy–galaxy lensing estimators and their covariance properties. *Monthly Notices of the Royal Astronomical Society*, 471(4):3827–3844, 07 2017. ISSN 0035-8711. doi: 10.1093/mnras/stx1828. URL https://doi.org/10.1093/mnras/stx1828.

Somerville, R. S. and Davé, R. Physical Models of Galaxy Formation in a Cosmological Framework. *Annual Review of Astronomy and Astrophysics*, 53:51–113, August 2015. doi: 10.1146/annurev-astro-082812-140951.

Springel, V. et al. First results from the IllustrisTNG simulations: matter and galaxy clustering. *Mon. Not. Roy. Astron. Soc.*, 475(1):676–698, 2018. doi: 10.1093/mnras/stx3304.

Verma, N., Boyer, E., and Verbeek, J. Dynamic filters in graph convolutional networks. *ArXiv*, abs/1706.05206, 2017.

Villanueva-Domingo, P., Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., Marinacci, F., Spergel, D. N., Hernquist, L., Vogelsberger, M., Dave, R., and Narayanan, D. Inferring halo masses with Graph Neural Networks. *arXiv e-prints*, art. arXiv:2111.08683, November 2021.

Vogelsberger, M., Marinacci, F., Torrey, P., and Puchwein, E. Cosmological simulations of galaxy formation. *Nature Reviews Physics*, 2(1):42–66, January 2020. doi: 10.1038/s42254-019-0127-2.

Zhang, X., Wang, Y., Zhang, W., Sun, Y., He, S., Contardo, G., Villaescusa-Navarro, F., and Ho, S. From Dark Matter to Galaxies with Convolutional Networks. *arXiv e-prints*, art. arXiv:1902.05965, February 2019.