Mocking The Universe Using Machine Learning

Harry George Chittenden Supervisor: Dr. Rita Tojeiro
University of St. Andrews

Motivation

The physical connection between the evolutionary history of a galaxy and the dark matter halo which hosts it can be modelled, using hydrodynamical simulations, abundance matching, etc (see fig. I). but lacks a general formulism [1].

Computing baryonic physics in line with N-body dark matter simulations adds complexity to the task, and limits the volume and resolution of simulations (i.e. the large and small scales of the universe) which can be captured [1,2].

The large-scale cosmic structures housing the largest galaxies, as well as small-scale local environments, can have a profound influence on the galaxy population [1,3]; we seek to include these simultaneously.

We use a machine learning algorithm, specifically, recurrent neural networks, to compose a model which can predict the stellar formation histories of galaxies from the assembly of their dark matter halos.

We hope to apply this methodology to large volume simulations, reproducing galaxy formation in fine detail, with accelerated computational efficiency.

Our predictions may be used to infer observational data (e.g. galaxy spectra) and create mock catalogues, to complement the DESI survey.

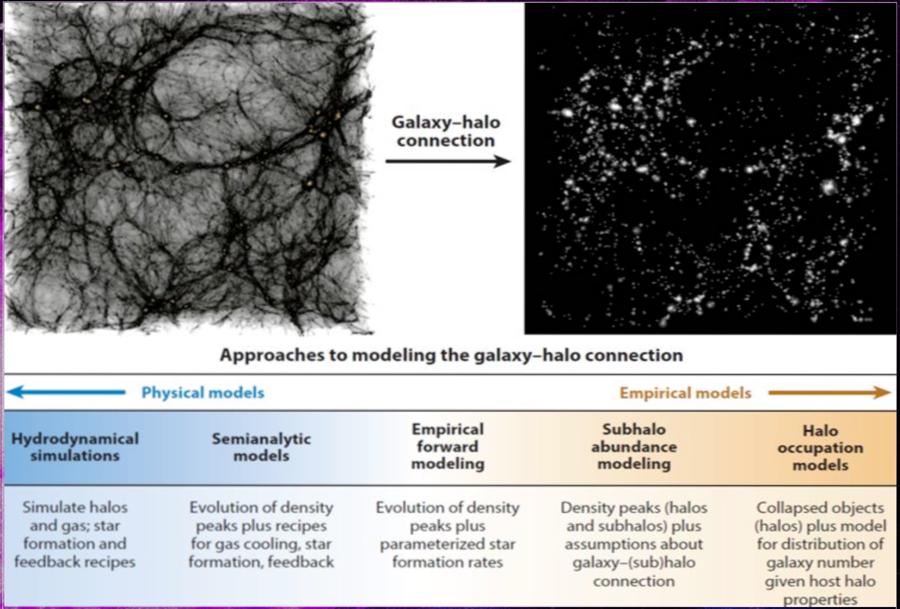


Fig. I: *A visual summary of existing measures used to model the galaxy-halo connection. The left image shows the dark matter distribution from an N-body simulation, to which a baryonic component can be applied by physical modelling. The right shows an empirical approach, in which galaxies of a given makeup are hierarchically assigned to halos based on observations [1].*

References
[1] Wechsler & Tinker, 2018, ARA&A, 56, 435
[2] Vogelsberger et. al., 2020, Nat. Rev. Phys., 2, 42
[3] Jo & Kim, 2019, MNRAS, 489
[4] Nelson et. al., 2019, Comp. Astro., 6
[5] Hani et. al., 2020, MNRAS, 493
[6] Engler et. al., 2020, MNRAS, 500
[7] Almeida et. al., 2012, ApJ, 756

Approach

The machine learning algorithm is being developed and tested using data from the IllustrisTNG cosmohydrodynamical simulations. We use TNG100-1 and TNG300-1: high mass resolution simulations in volumes $(100\text{Mpc})^3$ and $(300\text{Mpc})^3$, respectively.

In the TNG simulations, the larger volume simulations contain more objects and offer a better sampling of the highest mass galaxies. However, they have a much lower mass resolution [4]. We therefore use TNG100-1 to sample intermediate mass objects, and TNG300-1 for high mass objects.

Our neural networks use time-dependent and time-independent features; we are using a simple recurrent neural network layer for halo formation histories and standard dense layers for other properties.

Current Input Data: Halo mass at $z=0$, times at which fractions of the final mass formed, NFW concentration, halo mass accretion rates for all time.

Output Data: Galaxy Mass, stellar age spectrum.

In future, the local environment surrounding the target halos will be quantified and tested, and additional baryonic properties required to recreate observed data will be added.

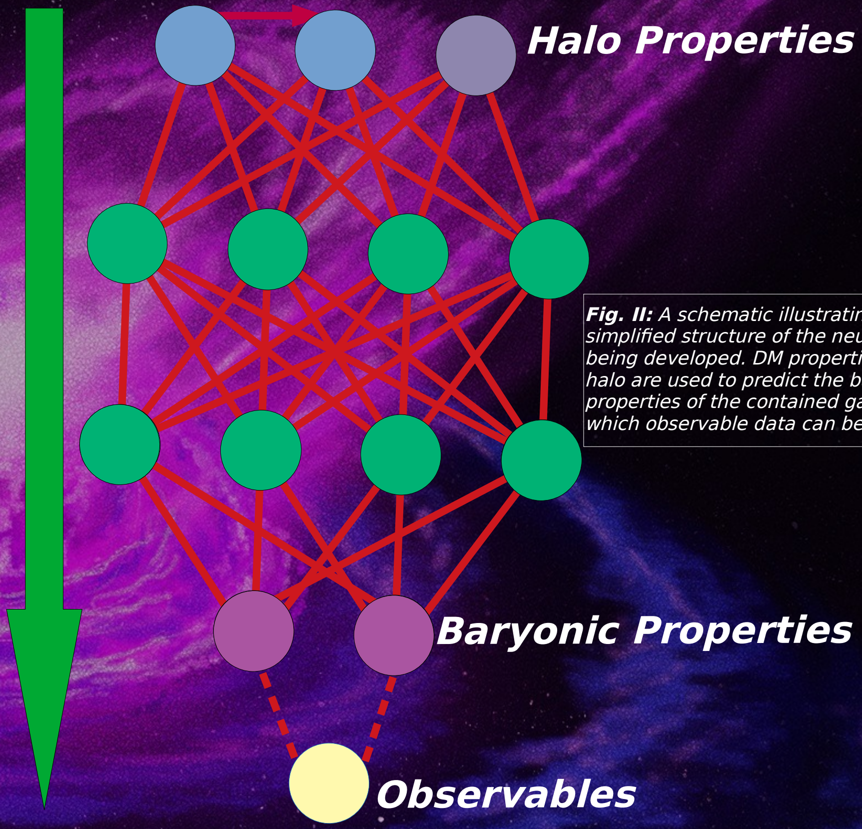


Fig. II: *A schematic illustrating the simplified structure of the neural network being developed. DM properties of the host halo are used to predict the baryonic properties of the contained galaxy, from which observable data can be inferred.*

References
[1] Wechsler & Tinker, 2018, ARA&A, 56, 435
[2] Vogelsberger et. al., 2020, Nat. Rev. Phys., 2, 42
[3] Jo & Kim, 2019, MNRAS, 489
[4] Nelson et. al., 2019, Comp. Astro., 6
[5] Hani et. al., 2020, MNRAS, 493
[6] Engler et. al., 2020, MNRAS, 500
[7] Almeida et. al., 2012, ApJ, 756

Results

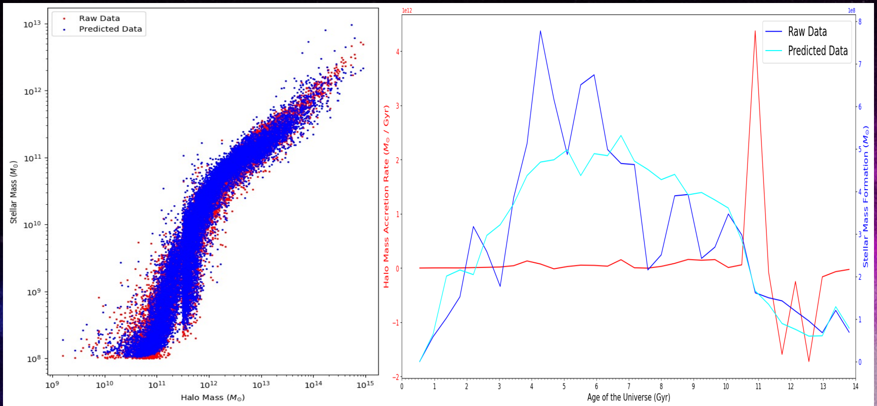


Fig. III: *The predictions of an example neural network. On the left, the relation between stellar and halo mass, where the predicted points (blue) closely match the trend and scatter of the original data (red). On the right, the halo assembly rate (red, $M_\odot$/Gyr) and the corresponding stellar mass assembly from the original data (blue) and predicted by the network (cyan). The trend in stellar assembly rate is well predicted; more subtle features are uncorrelated with the input data and therefore not predicted.*

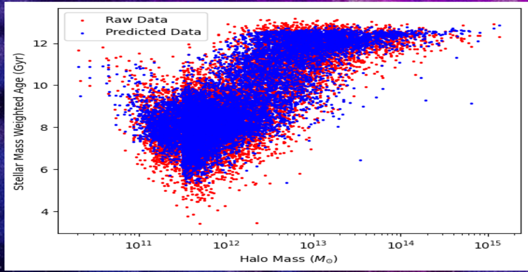


Fig. IV: *The mass-weighted stellar age of all target galaxies, plotted as a function of halo mass at z=0, and computed by averaging over the stellar age spectrum of the TNG simulation data (red) and the stellar age spectrum returned by the neural network (blue). The similar shapes of both distributions indicate that the stellar formation history has been well recovered.*

We have predicted the expected relation between stellar and halo mass, including the mass-dependent scatter, thereby capturing the variance in galaxy formation histories (see fig. III).

The mass-weighted stellar age of the galaxy is used to characterise the full stellar formation history and assess the credibility of the predicted age spectrum. The similar distributions of the true and predicted stellar ages in fig. IV show that these are in general agreement.

The smooth shape of the stellar accretion rate is apparent, but sharper features are mostly absent. On close inspection, many small features coincide with small variations in the halo accretion rate.

Larger, sharper changes in the stellar age spectra (e.g. spikes corresponding to star formation bursts) cannot be inferred from mass assembly alone [3,5], and are seldom shown by our neural networks.

We find that halo masses, early accretion and maximum accretion rates are strongly correlated with our targets. However, times of formation and merger information have been less informative.

We believe that the dark matter distribution in proximity to the target halo will be useful for predicting galaxy formation [3, 6], yet we have not tested this.

In short: AI is working, but it needs more information!

References

[1] Wechsler & Tinker, 2018, ARA&A, 56, 435
[2] Vogelsberger et. al., 2020, Nat. Rev. Phys., 2, 42
[3] Jo & Kim, 2019, MNRAS, 489
[4] Nelson et. al., 2019, Comp. Astro., 6
[5] Hani et. al., 2020, MNRAS, 493
[6] Engler et. al., 2020, MNRAS, 500
[7] Almeida et. al., 2012, ApJ, 756

Outlook

In future, we will attempt to characterise the local environment surrounding the halo and its merger history, as these have profound consequences on galaxy evolution [3,5]. Such quantities are alluded to in fig. V.

We will also develop this algorithm to predict corresponding star populations, and properties such as sSFR, age and metallicity; ultimately, the necessary information to construct galaxy spectra [7].

We will simulate a mock galaxy survey by applying the final neural network to a DESI-volume N-body dark matter simulation. This will capture the full range of halo masses and environments and reproduce a sample of galaxies much larger than the TNG simulations.

In order to identify the driving factors of the galaxy-halo connection, we may use covariance calculations or PCA (already used in some of our neural networks) to identify which halo and galaxy parameters are most tightly correlated. The mock surveys we construct should echo these constraints.

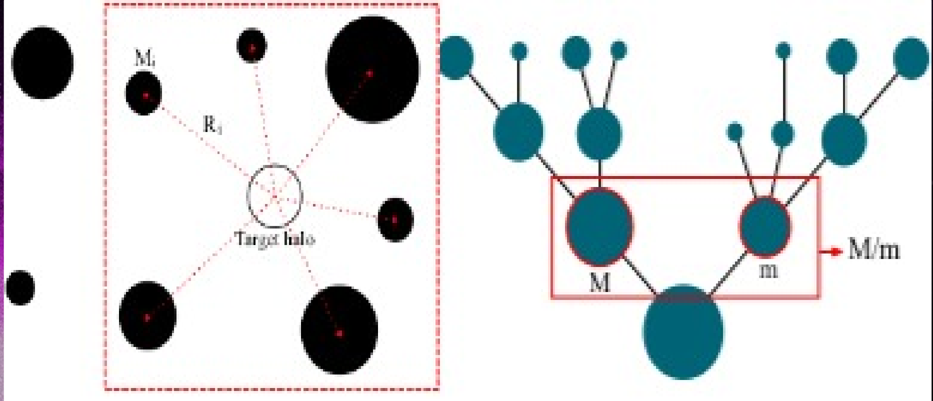
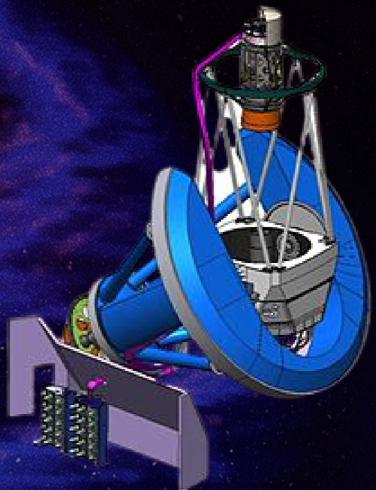


Fig. V: *A schematic showing how the local environment and merger history of a halo may be quantified. The left indicates the sum of mass over distance for all halos in proximity to a target object. The right depicts the mass ratio of the most recent merger event in a halo's history; applicable to all mergers at all times [3].*



References

[1] Wechsler & Tinker, 2018, ARA&A, 56, 435
[2] Vogelsberger et. al., 2020, Nat. Rev. Phys., 2, 42
[3] Jo & Kim, 2019, MNRAS, 489
[4] Nelson et. al., 2019, Comp. Astro., 6
[5] Hani et. al., 2020, MNRAS, 493
[6] Engler et. al., 2020, MNRAS, 500
[7] Almeida et. al., 2012, ApJ, 756