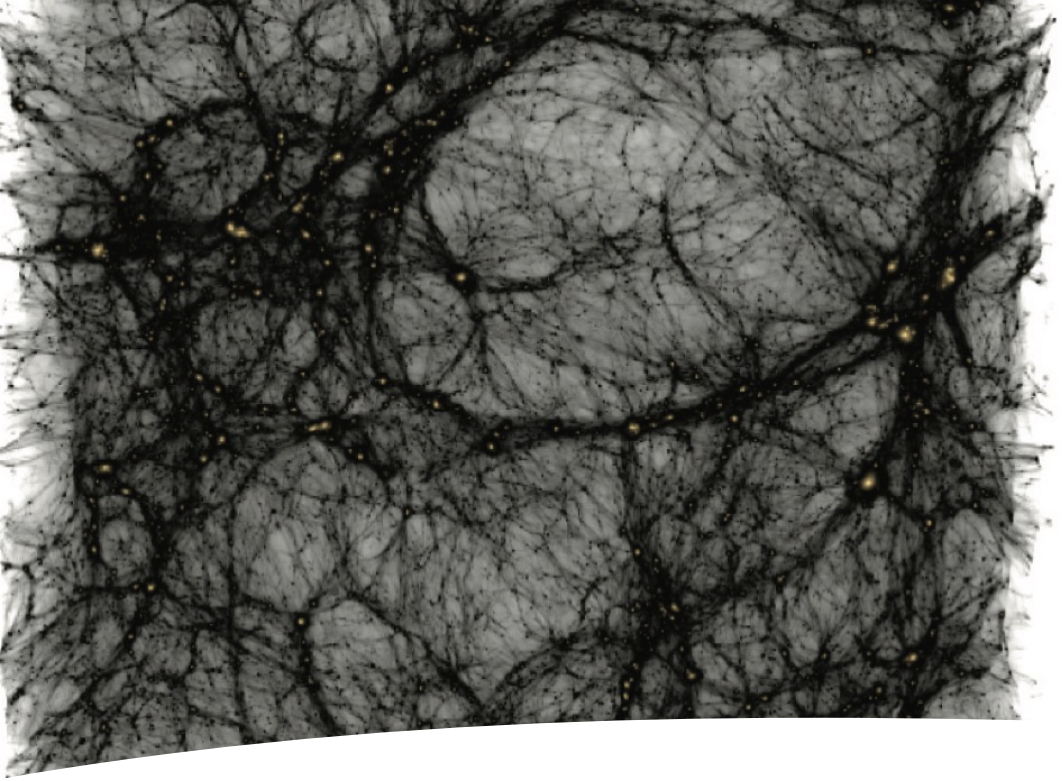
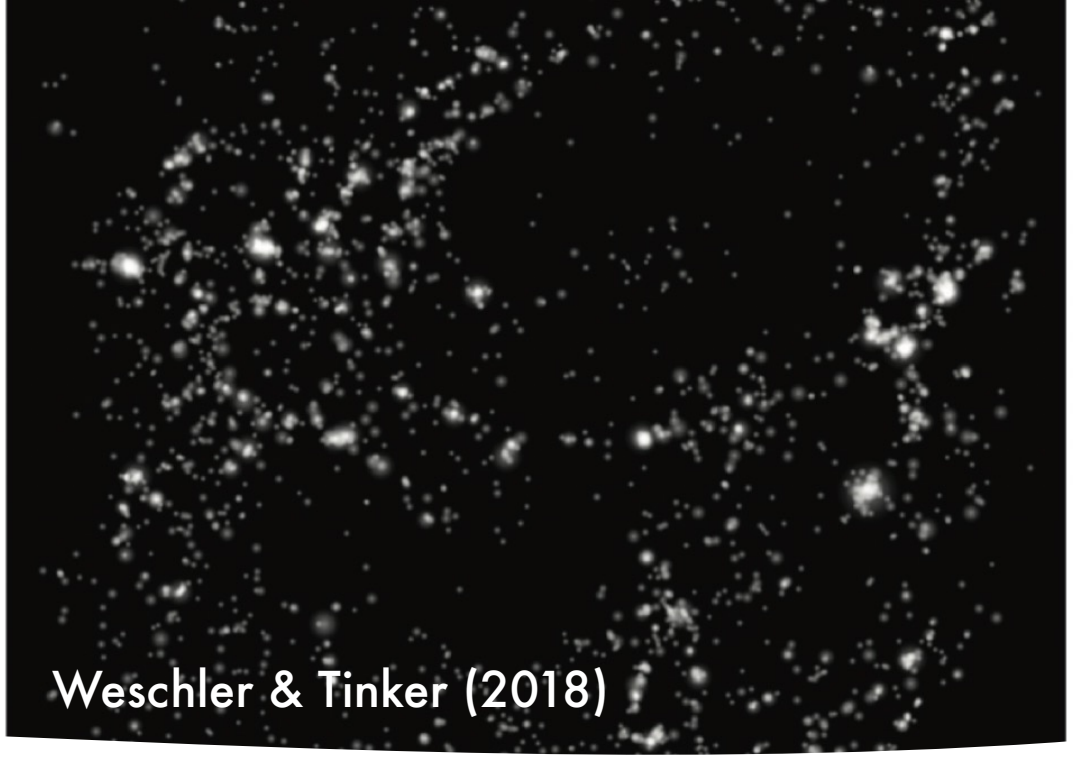
Constraining the Complex Relationships Between Galaxies and their Dark Matter Haloes with Machine Learning

Ryan Roberts



Galaxy–halo connection

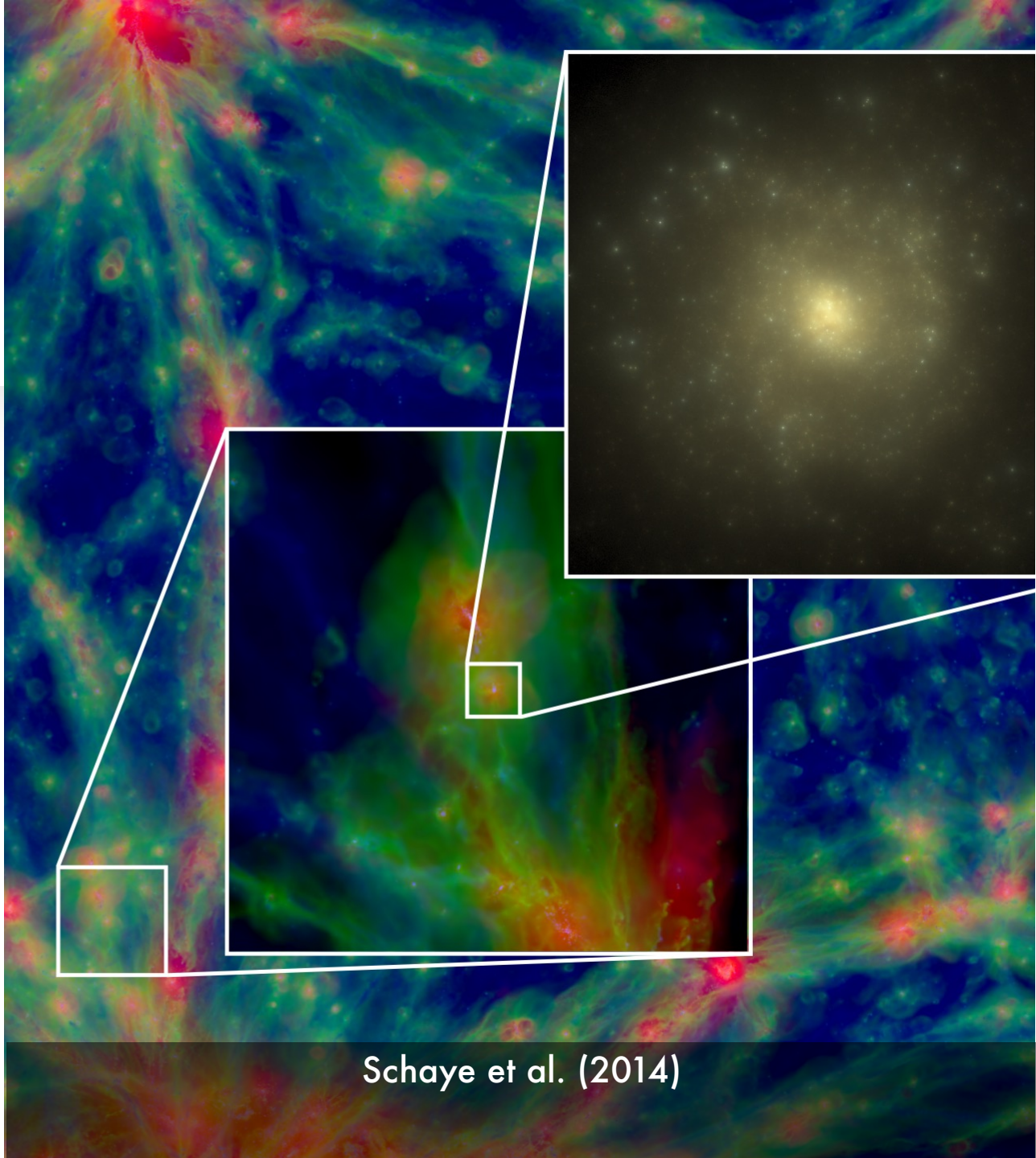


Cosmology and the Role of Dark Matter

- ΛCDM Universe → existence of cold dark matter.
- Galaxy surveys have shown us that galaxy population is very diverse.
- Galaxies form inside dark matter haloes. Simulations indicate that the properties of galaxies are highly correlated with features of the host halo – such as halo mass.
- However, many other properties of dark matter haloes have the potential to be correlated with certain galaxy properties (M_*, f_{CGM}). This behaviour is inferred from the scatter observed in galaxy properties at fixed mass.

Cosmological Hydrodynamical Simulations

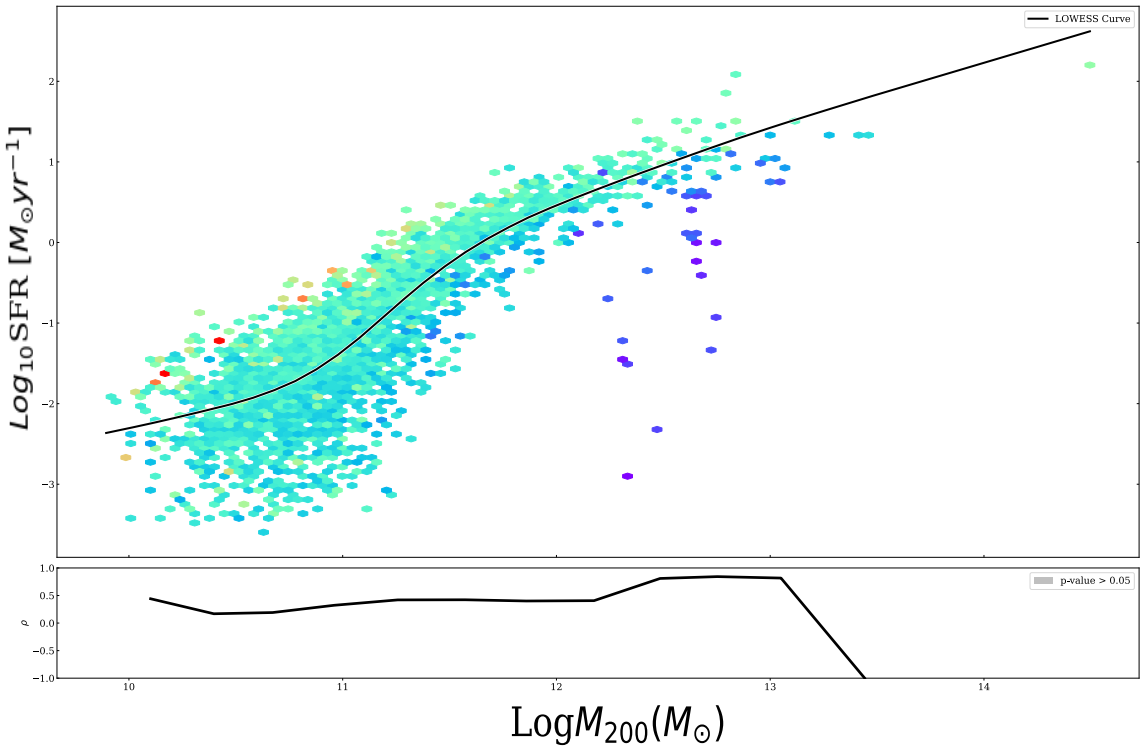
- Cosmology is different from most sciences; it cannot be directly experimented with → need for simulations.
- Hydro simulations allow astronomers to experiment and compare with observations.
- We can convincingly replicate a ΛCDM Universe and even alter initial conditions and cosmological parameters.

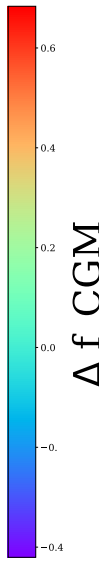


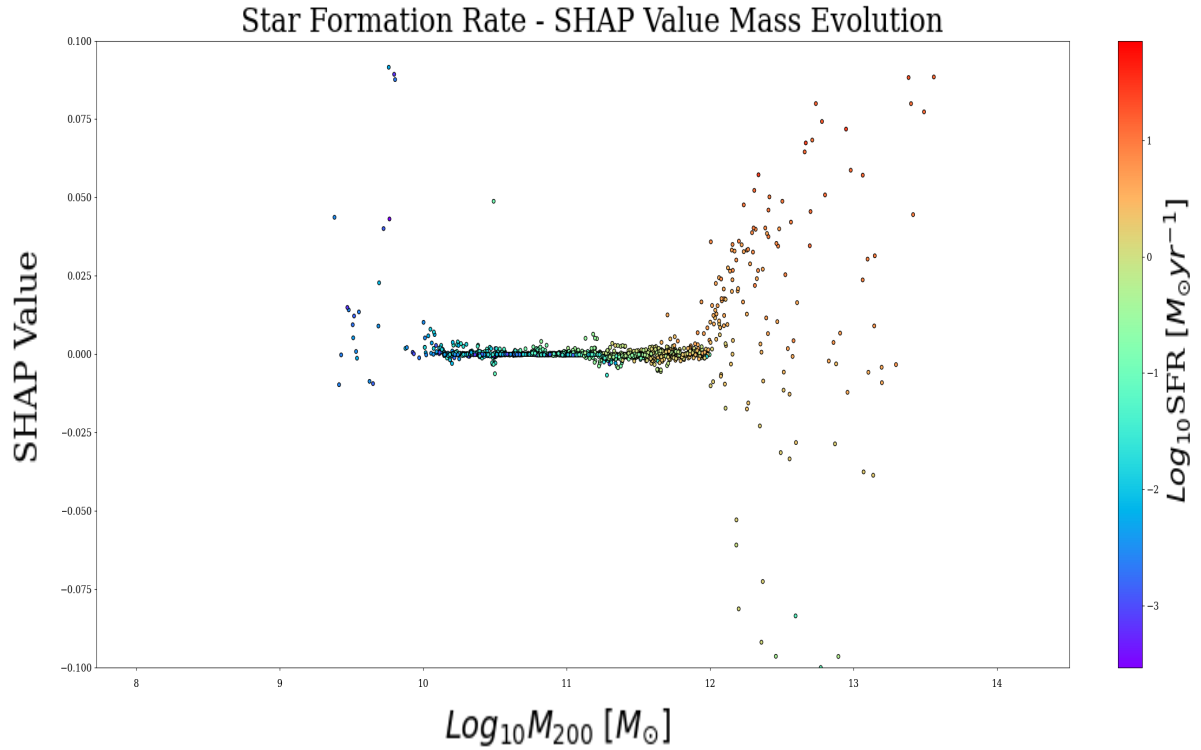
Schaye et al. (2014)

XGBoost and Feature Importances

- The halo population itself is very diverse at fixed mass – so many factors can contribute to the observed diversity of galaxies. Since these are generally correlated, we require tools to help quantify covariances in properties.
- XGBoost is a form of ensemble predictor; it provides an extra tier of capability by feeding previous tree decisions into new trees.
- Ensemble predictors are inherently straightforward to garner feature importances from → crucial for project.
- Example: Determine which parameters contribute most towards predicting f_{CGM} in a given mass range. A plot of the SFR against halo mass (coloured by the change in f_{CGM}) shows that at the highest mass values, the low values of SFR are attributed to the greatest decrease in f_{CGM}. Theoretically, at high masses you would expect to see the highest SFRs at the highest values of f_{CGM}, as there is more material that can lead to star formation.







Future Direction / Explainable AI

- Feature importance plots are promising, but how trustworthy and explainable are SHAP and other feature importance methods?
- Could another ML method be useful for this kind of challenge?

Thank you!