**Inferring galaxy dark halo properties from visible matter with Machine Learning**

**Link:** [**https://arxiv.org/abs/2111.01185?context=astro-ph**](https://arxiv.org/abs/2111.01185?context=astro-ph)

**Original abstract:**

Next-generation surveys will provide photometric and spectroscopic data of millions to billions of galaxies with unprecedented precision. This offers a unique chance to improve our understanding of the galaxy evolution and the unresolved nature of dark matter (DM). At galaxy scales, the density distribution of DM is strongly affected by the astrophysical feedback processes, which are difficult to fully account for in classical techniques to derive mass models. In this work, we explore the capability of supervised learning algorithms to predict the DM content of galaxies from “luminous” observational-like parameters, using the public catalog of the TNG100 simulation. In particular, we use, Photometric (magnitudes in different bands), Structural (the stellar half-mass radius and three different baryonic masses) and Kinematic (1D velocity dispersion of all particles and the maximum rotation velocity) parameters to predict the total DM mass, DM half-mass radius, DM mass inside one and two stellar half-mass radii. We adopt the coefficient of determination, 𝑅2, as a reference metric to evaluate the accuracy of these predictions, whose values are between 0 (worst fit possible) and 1 (perfect match). We find that the Photometric features alone are able to predict the total DM mass with fair accuracy (𝑅2 􏰀 0.86), while Structural and Photometric features together are more effective to determine the DM inside the stellar half mass radius (0.88 􏰀 𝑅2 􏰀 0.94), and the DM within twice the stellar half mass radius (0.90 􏰀 𝑅2 􏰀 0.92). However, using all observational quantities together (Photometry, Structural and Kinematics) incredibly improves the overall accuracy for all DM quantities (up to 𝑅2 ∼ 0.98). This first test shows that Machine Learning tools are promising approaches to derive predictions of the DM in real galaxies. The next steps will be to improve observational realism of the training sets, by closely select samples which accurately reproduce the typical observed “luminous” scaling relations. The so-trained pipelines will be suitable for real galaxy data collected from the next-generation surveys like Rubin/LSST, Euclid, CSST, 4MOST, DESI, to derive, e.g., the properties of their central DM fractions.

**Re-written abstract:**

Next-generation surveys (including Rubin, Euclid, CSST, 4MOST, and DESI) will furnish cutting-edge data on an extensive sample of galaxies. These state-of-the-art photometric and spectroscopic datasets will offer an opportunity to further our understanding of galaxy evolution and dark matter (DM). However, the density distribution of DM at galaxy scales is heavily influenced by astrophysical feedback processes. These complicated processes are difficult to fully incorporate into classical techniques to constrain DM masses at these scales. We propose to expand beyond classical mass models with a supervised learning algorithm to infer the DM content of galaxies from luminous matter using data from the TNG100 simulation public catalog. We will adopt several observable data metrics, including brightness in different bands, baryonic masses, stellar half-mass radii, and velocity dispersions. Using these parameters, we will recover DM masses and radii with verifiable accuracies from the TNG100 simulation. Future applications of this pipeline to next-generation surveys will allow for critical predictions about DM from the real galaxy data observed.