MACHINE LEARNING FOR ASTROPHYSICS
2ND EDITION CATANIA, 8-12 JULY, 2024

Contribution ID: **184** Type: **Oral Presentation**

Inferring Galaxy Baryonic Properties from IllustrisTNG Dark Matter Merger Trees with Graph Neural Networks

Wednesday, 10 July 2024 09:40 (20 minutes)

As more and more data from recent and upcoming large cosmological surveys become available, the need for equally detailed theoretical models of galaxy formation, including large-volume cosmological simulations emerges. The complexity of such simulations, however, implies a great computational cost, which to this day always leads to a compromise in either resolution or size. As a supplement to cosmological simulations, we perform supervised learning to create a mapping between the baryonic and the dark matter (DM) component in high-resolution hydrodynamical simulations. We use the merger trees from TNG100 of the IllustrisTNG simulation suite, that represent the formation history of DM subhalos in the form of graphs, along with Graph Neural Networks (GNNs), which excel in handling datasets that carry such a distinct intrinsic structure as graphs. We train a GNN to infer galaxy baryonic properties, such as stellar and gas mass, given the host subhalo DM properties across the merger tree. We obtain a potent model that is capable of reproducing galaxy populations that are statistically consistent with simulated ones. This powerful tool can generate entire galaxy populations for cosmological volumes: one can run computationally inexpensive, simple DM-only (N-body) simulations and subsequently augment them, by "painting"galaxies on top of the DM subhalos, enabling the construction of Gpc-scale boxes with baryonic physics. In order to tackle the vastly diverse –in size and complexity–graphs that comprise the merger trees, we utilize a specific class of GNNs called Graph Attention Networks (GATs), which also involve additional attention coefficients that capture the importance of certain features in the graphs to determining the desired outputs. These coefficients also provide a degree of interpretability, since they correspond to actual events in the formation history of the subhalos, and can thus be studied to extract information about the interplay between galaxies and their hosts'assembly history. This method serves not only as a proof of concept, that the model is not just a "black box", rather it learns physically meaningful representations, but also as a scientific tool that can be used to further understand the complex galaxy-halo connection.

Presenter: ANDREADIS, Nick (Astronomical Observatory, Ghent University)

Session Classification: Cosmology & Simulations