Deep learning for dark matter and dark sirens

Cosmological inference from the large-scale structure of the Universe is entering an exciting decade with an unprecedented wealth of data coming from new surveys like DESI (the largest spectroscopic survey ever undertaken), the ESA Euclid mission, and the Rubin Observatory in each of which UCL has major stakes. Traditional analytic models of the key cosmological signals, such as gravitational lensing and galaxy clustering, increasingly struggle to cope with the complexity and precision of new datasets. Therefore, we are faced with several difficult inverse problems of processes that can only be quantified through computationally expensive numerical simulations of the large-scale dark matter distribution, of how galaxies assemble within this dark matter scaffolding, and of the galaxy light's journey through the cosmos into our telescopes.

The supervisory team have pioneered a simulation-based inference approach via neural density estimation that can directly infer properties of dark matter and other cosmological parameters from suites of comprehensive simulations, many of them created in-house. The approach combines data compression, Bayesian inference, and state-of-the-art machine learning techniques like normalising flows. The student will further develop the inference technique towards competitive speed and accuracy for the new generation of surveys like Euclid, typically an order of magnitude larger than current datasets.

An exciting new way to access information from the deep Universe beyond electromagnetic radiation is rapidly maturing: gravitational waves. A single gravitational wave event with an electromagnetic counterpart observed in 2017 already put important constraints on our cosmological model. Orders of magnitude more gravitational wave signatures are now detected without an electromagnetic complement. With more sophisticated inference techniques, these dark sirens can still be used to infer cosmological model parameters and constrain extensions of general relativity.

The analysis will again be undertaken in a simulation-based inference framework. The student will build upon existing cosmological simulation code developed at UCL to generate realistic mock data given an input model: this can improve over existing dark siren analyses in terms of reliability and discovery potential. A significant aspect of the project will be performing optimal feature extraction/data compression using deep learning methods. This will be an opportunity to explore geometric learning approaches, like deep sets and graph neural nets.

This PhD project will begin by further developing and optimising the methodology for the dark matter (Year 1) and dark siren (Year 2) inference, thereby building a deep understanding of the data, the underlying physics, and the machine learning and data science concepts we apply. Post-placement, the focus will be on real-data applications of the methods, engaging with ~1 billion galaxies observed in Euclid's second data release in combination with the latest overlapping data from DESI (electromagnetic spectra) and Ligo/Virgo/Kagra (gravitational waves) in Years 3 and 4.

The project will be supervised jointly by Prof Benjamin Joachimi and Dr Niall Jeffrey. Joachimi is an international expert in the analysis of cosmic large-scale structure data, with leadership roles in leading galaxy surveys (Euclid, Rubin Observatory, ESO Kilo-Degree Survey). He has an established track record of translating data science concepts into novel astrophysical applications (including data compression, uncertainty quantification, saliency mapping, and Bayesian inference) and is a member of the CDT-DIS management team. Jeffrey is an expert on developing deep learning-based approaches to inference. He has a leading role in the production of dark matter sky maps and simulation-based inference with optimal feature extraction in the Dark Energy Survey and was granted time on the UK's national high-performance computing infrastructure to create suites of cosmological simulations optimised for inference.