Tristan Deleu

Causal discovery with Generative Flow Networks

Discovering the structure of a causal model purely from data is plagued with problems of identifiability. In general, shy of any assumptions about how the data was generated, multiple equivalent models may explain our observations equally well even if they could entail widely different causal conclusions. As a consequence, choosing an arbitrary element among these equivalent models may result in making unsafe decisions if it is not aligned with how the world truly works. It is therefore essential to keep a notion of epistemic uncertainty about our possible candidates in order to mitigate the risks posed by these misaligned models, especially when the data is limited. In this talk, I will introduce a new class of probabilistic models called Generative Flow Networks (GFlowNets) that provides a general framework for modeling probability distributions over discrete and compositional objects, such as the structure of a causal model, which is represented as a directed acyclic graph (DAG). GFlowNets treat generation as a sequential decision making problem, where each sample is constructed piece by piece. I will highlight how they connect to various domains of machine learning and statistics, including variational inference and reinforcement learning. Finally, I will show how GFlowNets can be used for causal discovery from a Bayesian perspective, to model the whole posterior distribution over causal models with arbitrary mechanisms, given a dataset of observations.