# All Terrain Graph-Learner-CPC: Causal Model Discovery for Missing Not at Random Data

**Abstract**:
Missing data, a prevalent problem in biological data, can compromise statistical inference, thus hindering downstream network analysis. Most existing methods to handle missing data assume how the data are missing is either independent from the observed variables or depends on some variables which do not have missing values. However, in biological data, such as scRNA-seq data, often how the data are missing is dependent on some variables which have missing values. This paradigm results in Missing Not at Random (MNAR) data. Traditional missing data methods have been shown to produce biased estimates for MNAR data, which are passed on to causal network analysis. We introduce all terrain graph-learner-CPC (ATG-CPC), a novel method to learn causal graphs from MNAR data. ATG-CPC uses all of the observed data to directly impute causal edges. The key to our approach is learning and applying a conditional independence test transformation to calculate sample size adjusted conditional independence p-values for every potential edge. Using simulated data we show ATG-CPC to be an effective tool for recovering causal graphs masked by MNAR values. ATG-CPC generates more accurate causal graphs than causal learning methods even after pre-processing the data with traditional missing data methods.