# BENCHMARKING A CAUSAL DISCOVERY METHOD FOR PARTIALLY OBSERVED BIOCHEMICAL KINETICS

H. Chambers[1], B. Bravi[1], V. Shahrezaei[1], H. Isambert[2]

[1] Department of Mathematics, Imperial College London, Huxley Building, 180 Queen's Gate Terrace, SW7 2AZ

[2] CNRS UMR168, Institut Curie, Université PSL, Sorbonne Université, 13 Rue Pierre et Marie Curie, 75005 Paris, France

Biological systems are inherently complex, often involving components that are difficult or impossible to measure. One effective way to represent these systems is through network models, where nodes correspond to variables and edges represent the interactions between them. Such network representations provide valuable insights that can support more advanced modelling efforts. Causal discovery methods uncover these interactions, revealing hidden effects in two primary ways: as common causes of observable variables and through time-lagged effects resulting from intermediate causes (Fig. 1).

In this study, we evaluate the performance of the temporal Multivariate Information-based Inductive Causation (tMIIC) method for causal discovery in the context of partial observations of time series data derived from biochemical kinetic models. Our results demonstrate tMIIC's high recall in uncovering the interactions among chemical species within these reaction systems. By omitting some data, we also show that tMIIC is most effective at detecting hidden nodes which appear as intermediate, time-lagged effects.

Our work shows the feasibility of causal discovery methods as a component in the wider mathematical modelling pipeline of biochemical reactions. By incorporating causal discovery into the modelling process, hidden structures and interactions may be uncovered that would previously have gone undetected. This enhances our understanding of the dynamics governing complex biological systems, allowing for more accurate and insightful predictions.

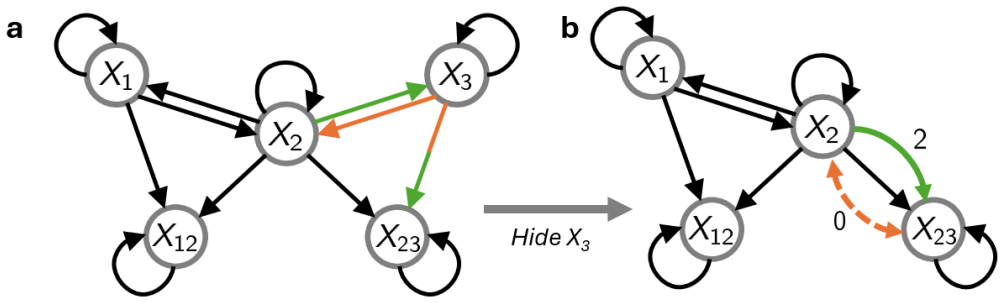


Fig. 1 - $X_3$ is a **common cause** of $X_2$ and $X_{23}$, and is an intermediate over a **time delayed** effect. **a**, the theoretical network from fully observed data. **b**, the theoretical network with $X_3$ unobserved.