

GNN-Causal-Inference

实验1: PCMCI+算法验证

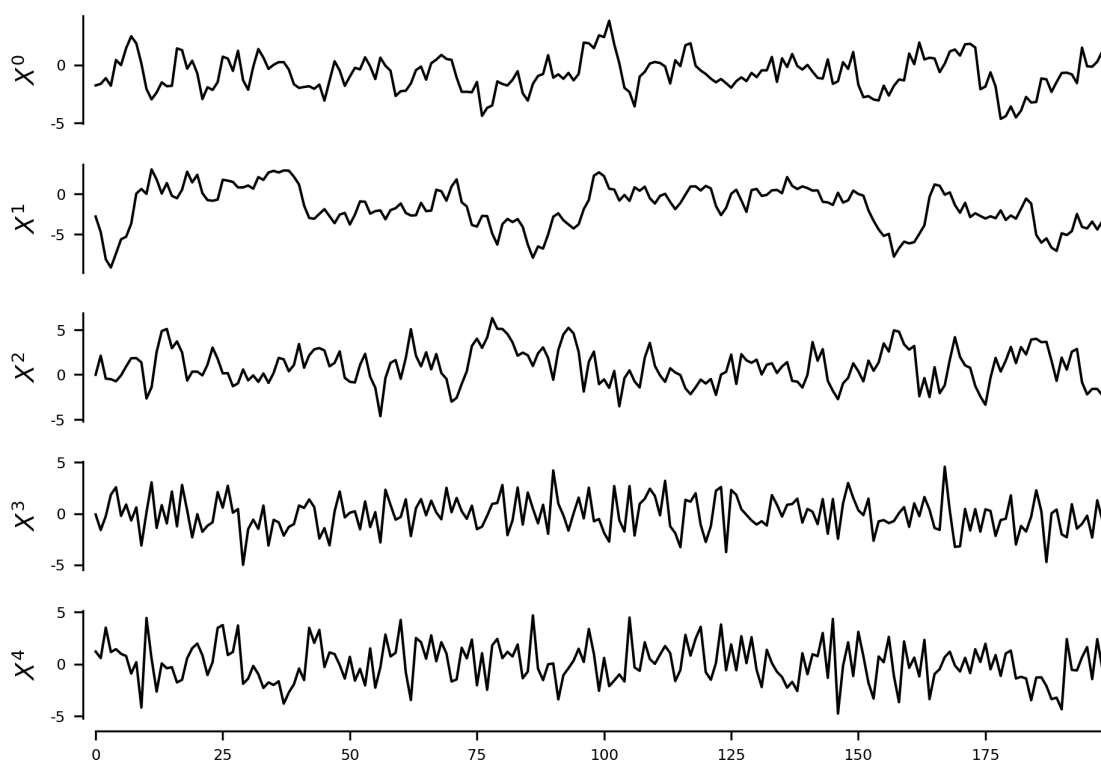
参考:

- <https://hekai.blog.csdn.net/>
- https://github.com/jakobrunge/tigramite/blob/master/tutorials/benchmarking_and_validation/tigramite_tutorial_explaining_correlations.ipynb

对比分析PCMCI+算法生成的因果图和显著相关图

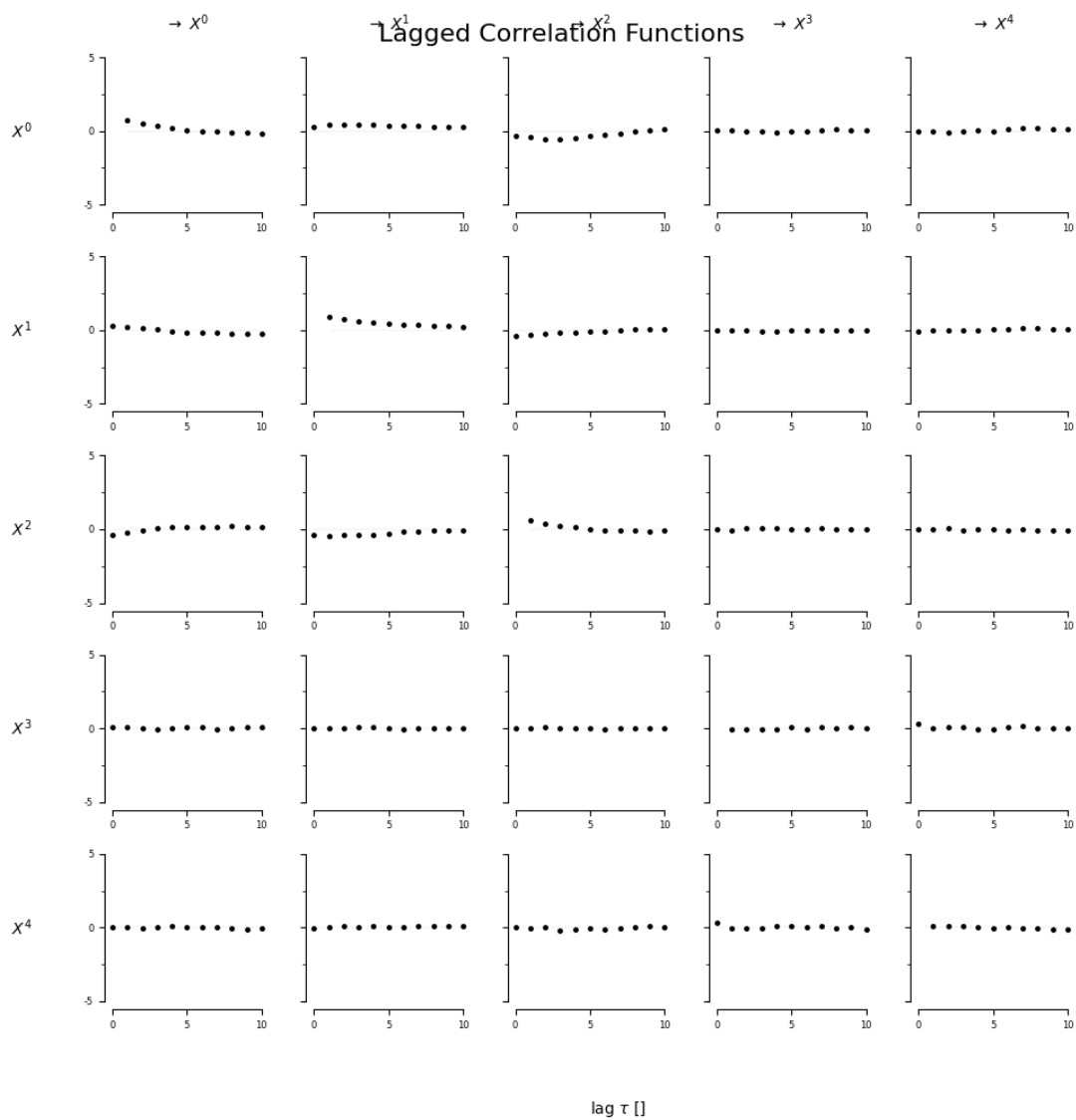
1. 数据合成: 根据 $X_t = \alpha \cdot Y_{t-1} + \beta \cdot Z_{t-2} + \epsilon_t$ 预设因果关系, 并添加噪声, 生成合成时间序列数据;

Observed Time Series Data



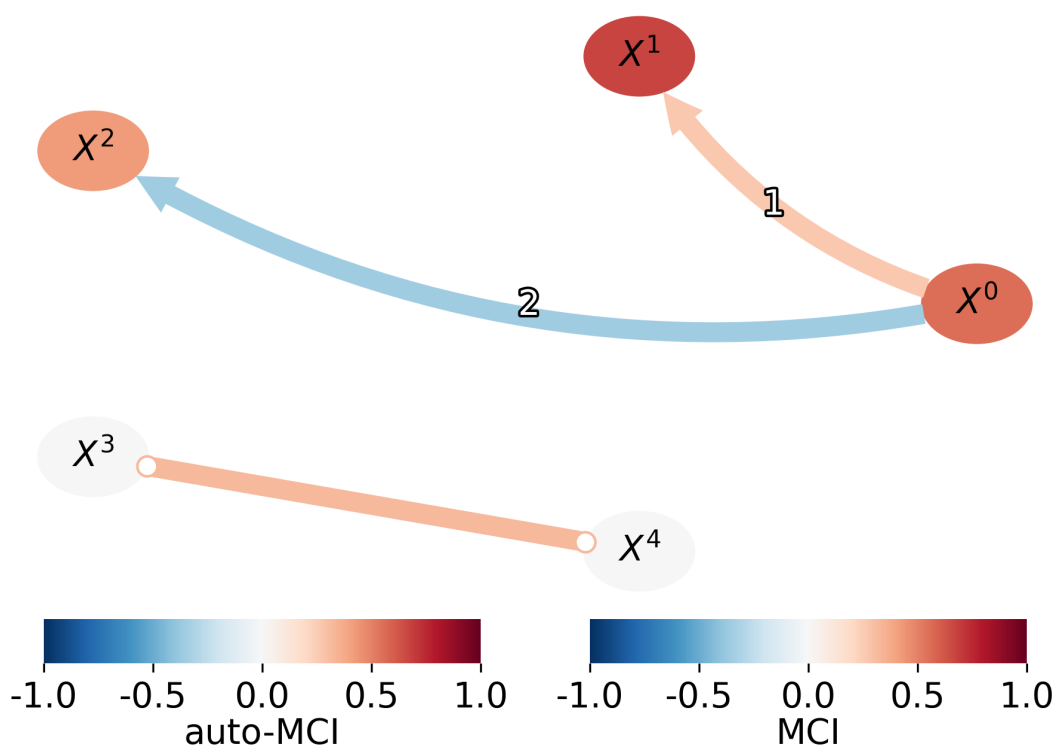
2. 计算滞后相关函数

观察变量之间是否存在时间延迟上的相关性,确实符合预设的因果关系



3. PCMCI+学习因果图：从观测数据中识别出：哪些变量对另一个变量有直接的滞后影响，这些影响是正向还是负向的（通过 MCI 数值判断）

Estimated Causal Graph using PCMCI+

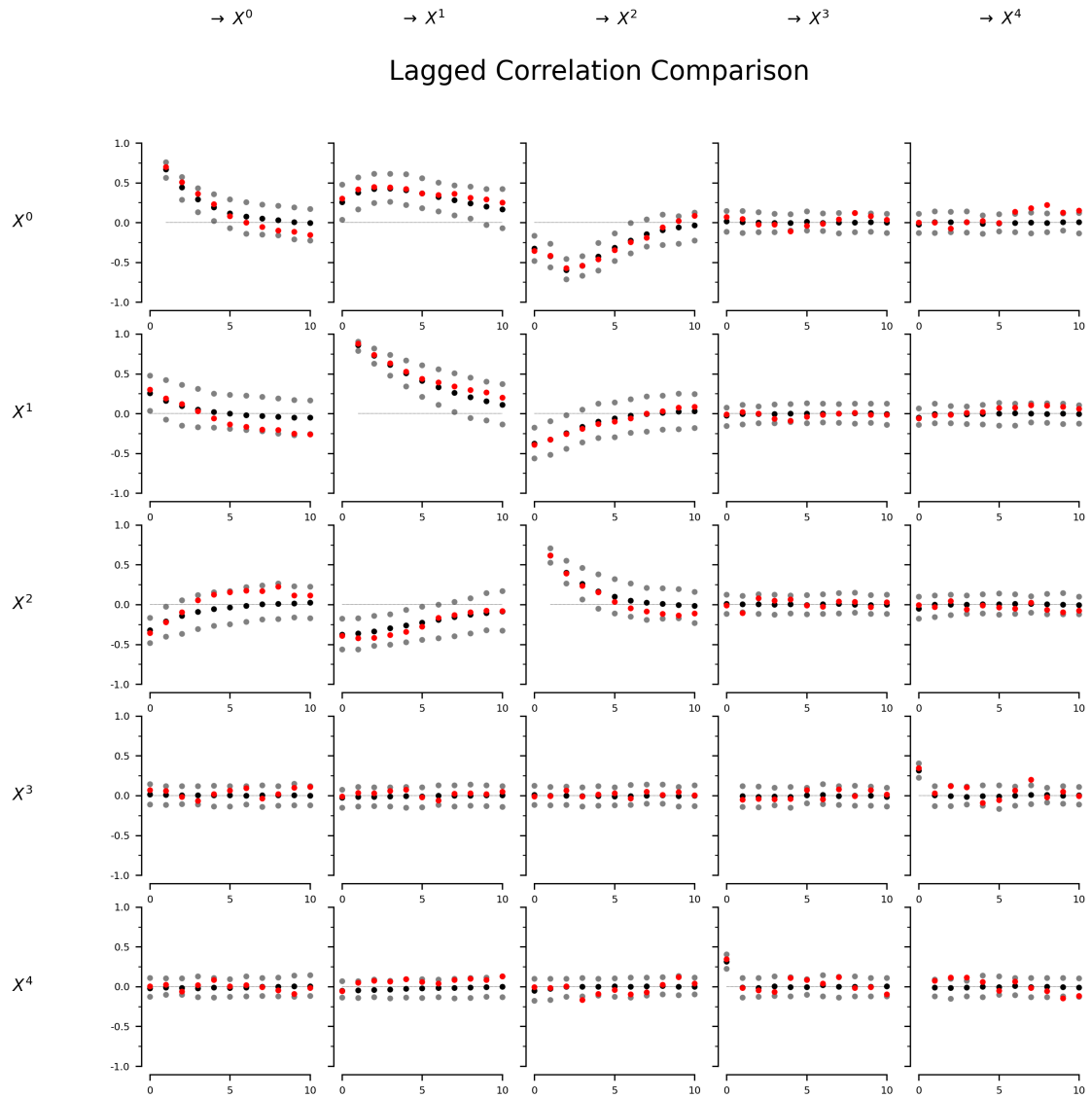


4. 构建显著相关图用于对比

- 1) 生成代理数据集：基于原始数据和已知因果结构，生成多个替代数据集（具有类似的统计特性、不同的随机噪声实现、“无因果”的参考分布）
- 2) 对比原始相关性和代理数据的相关性，判断其是否显著

相关性成因：

- 真实的因果关系；
- 样本量有限或噪声引起的伪相关性)



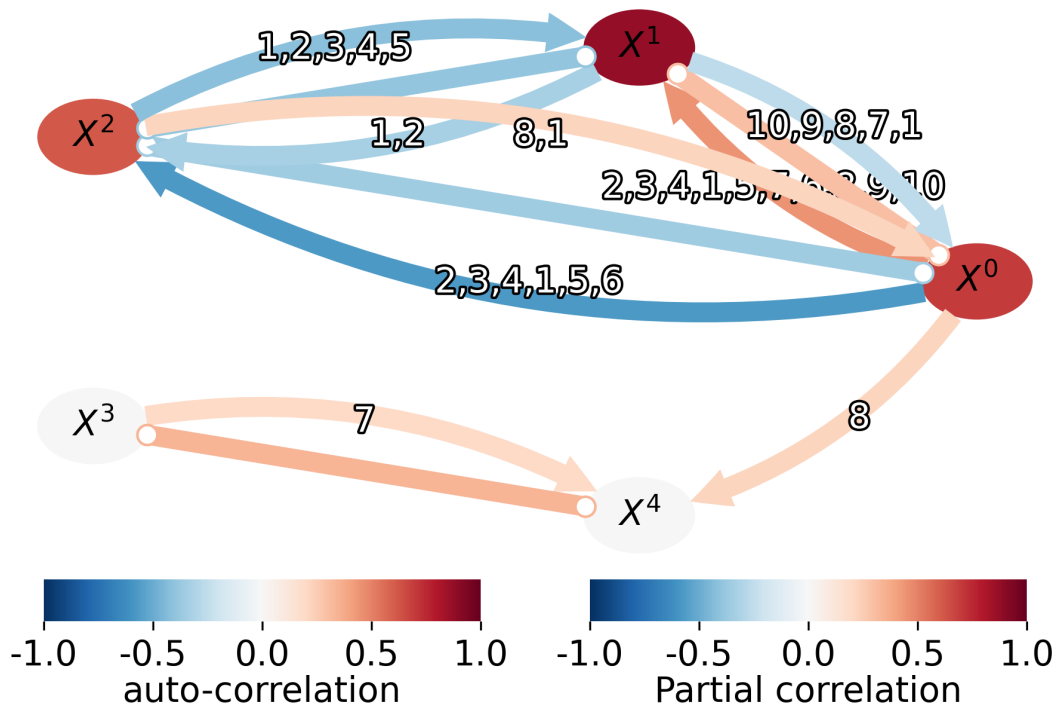
黑色线：替代数据的平均相关性

灰色区域：90%置信区间

红色线：原始数据的相关性

若原始相关性绝对值超出随机情况→存在真实的因果关系，否则表明原始数据中的相关性可能是随机波动的结果。

Significant correlation graph



筛选P值绘制显著相关图，显著相关图包含相关但非因果的边，但是显著性高的边确实为真是因果关系，PCMCI+算法的因果学习结果基本正确