

A GPU-Accelerated Skeleton Discovery for Gaussian Distribution Models

Motivation

The estimation of causal graphical models allows to solve important problems in many different domains. For example in genetics, gene regulatory networks can be seen as a practical embodiment of systems biology and can be used for drug design or diagnostics:

- Causal inference requires algorithmic support due to a rising complexity;
- Inefficiency of the common algorithms hinders its application in practice;
- Harness the processing power of the GPU to address the algorithm's computational complexity.

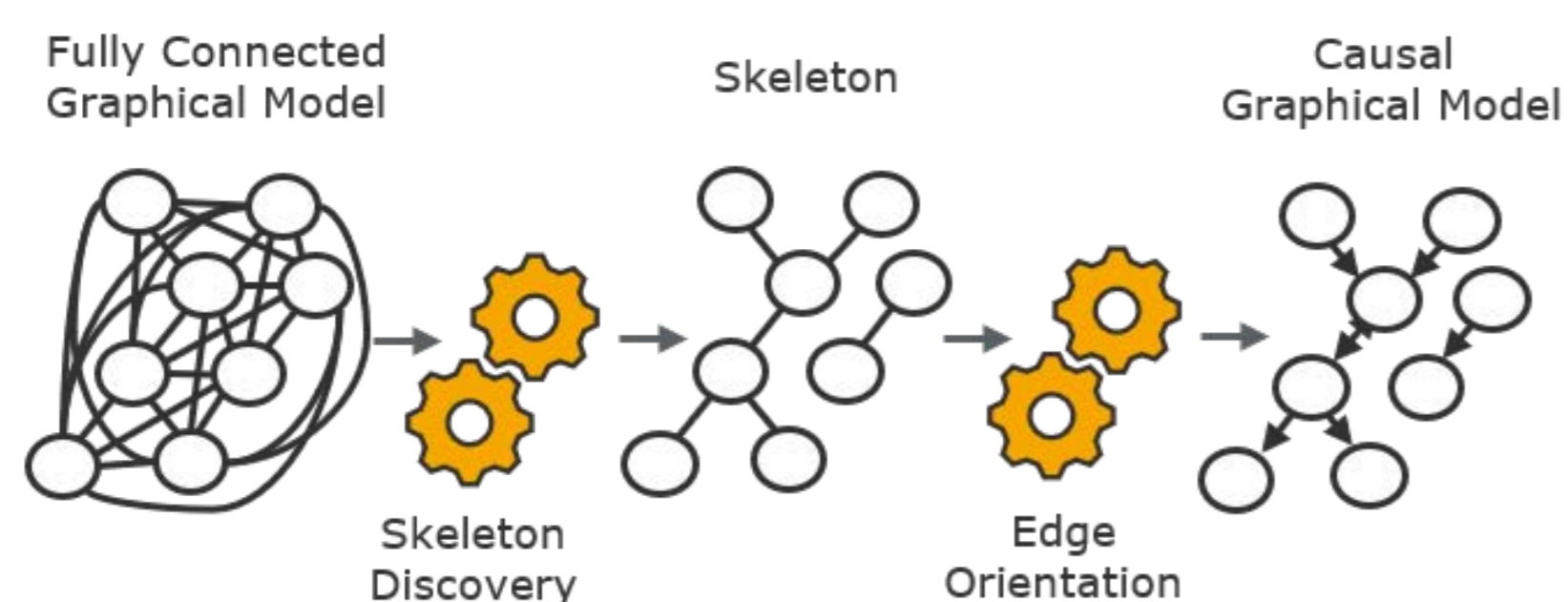


Figure: A schematic representation of the causal inference procedure

Causal Inference Procedure

In the recent years, the notion of causality has grown from a nebulous concept into a mathematical theory¹. A conceptual algorithm for learning the causal graphical model operates in two phases²:

- *Skeleton Discovery*, use conditional independence tests to receive information about the underlying relationships.
- *Edge Orientation*, determine the orientation of the detected relationships to construct a causal graphical model.

Preliminary Results

In our work, we investigate the use of GPU to address the high computational complexity of the causal inference and speed up its execution.

We propose a first version of a fully CUDA-based implementation of the skeleton discovery and provide an evaluation depicted in Table 1 based on real world gene expression data³.

The results indicate a benefit for lower levels, but also show the following limitations of our current version for higher levels:

- Low utilization of the Streaming Multiprocessors
- Large number of kernel calls to conduct a single conditional independence test

Based on these results, further improvements to fully utilize the processing capabilities of the GPU will be part of our future work.

	NCI-60	BR51
level l=0	53.75	71.6
level l=1	81.39	258.26
level l=2	0.004	0.003
level l=3	0.0002	0.0001
level l=4	0.001	0.0002
Total Including Data Transfer	0.48	0.68

Table: Average speed-up (of 10 runs) of GPU-accelerated against CPU-based skeleton discovery on two gene expression datasets

References:

- 1) J. Pearl. Causality: Models, Reasoning and Inference. Cambridge University Press, New York, NY, USA, 2nd edition, 2009.
- 2) P. Spirtes, C. N. Glymour, and R. Scheines. Causation, prediction, and search. MIT press, 2000.
- 3) T. D. Le, T. Hoang, J. Li, L. Liu, and H. Liu. A fast PC algorithm for high dimensional causal discovery with multi-core pcs. CoRR, abs/1502.02454, 2015.

Projektbeteiligte

Christopher Schmidt, Johannes Huegle, Siegfried Horschig and Matthias Uflacker
Hasso-Plattner-Institute
Enterprise Platform and Integration Concepts
August-Bebel-Str. 88
14482 Potsdam, Germany
{christopher.schmidt, johannes.huegle, matthias.uflacker}@hpi.de,
siegfried.horschig@student.hpi.de