

Learning causal relationships between genes from steady state data: algorithms, simulation and application

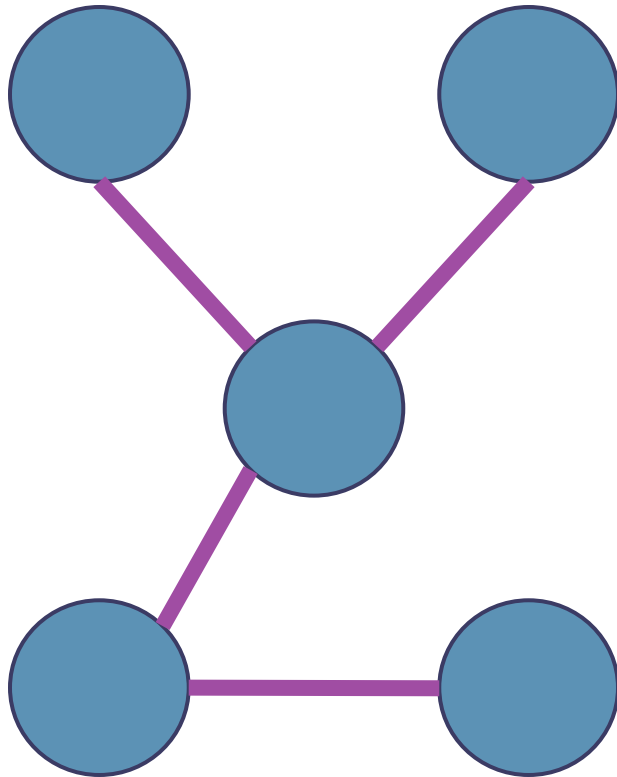
Michael Verdicchio, Xin Zhang, Seungchan Kim, Chitta Baral
Department of Computer Science and Engineering
Arizona State University
Tempe, Arizona, USA



Motivation

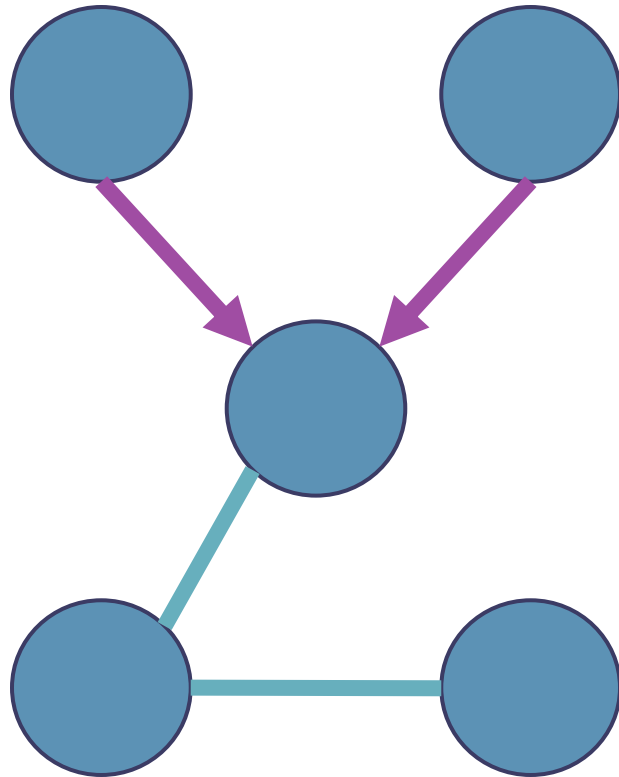
- Causal Bayesian Networks (BNs) have stronger implications than traditional BNs, and they facilitate intervention studies *in silico*
- Temporal data is helpful in learning, but could also be misleading
- Most accessible high-throughput human cancer data is in steady state form

Pearl's Inductive Causation



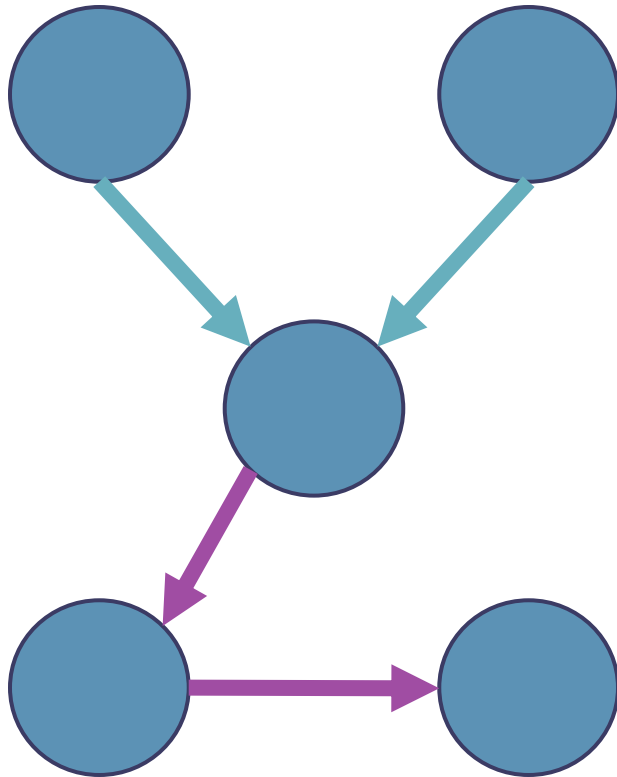
- Input: A data set
- Output: A Partially Directed Acyclic Graph (PDAG)
- Causal structure learned from steady-state data in steps
- **First the skeleton is learned**
- Then the edges are oriented where possible

Pearl's Inductive Causation



- Input: A data set
- Output: A Partially Directed Acyclic Graph (PDAG)
- Causal structure learned from steady-state data in steps
- First the skeleton is learned
- **Then the edges are oriented where possible**

Pearl's Inductive Causation



- Input: A data set
- Output: A Partially Directed Acyclic Graph (PDAG)
- Causal structure learned from steady-state data in steps
- First the skeleton is learned
- **Then the edges are oriented where possible**



The Modified IC (mIC) Algorithm

- Take advantage of genetic topological ordering sources – i.e. pathway databases
- The mIC algorithm only orients edges respecting prior topological ordering information
- A second algorithm (mIC_CoD) optimizes the number of causal parents using the coefficient of determination

Simulation and Application

Synthetic Data Sets

- mIC and mIC_CoD outperformed IC and traditional BN learning on small synthetic data sets

Melanoma Data Set

- mIC and mIC_CoD identified literature-verified causal relationships in a small melanoma data set

More details available on our poster!

Current and Future Work

- In-depth comparison of IC, mIC, mIC_CoD and BN learning
- Larger numbers of genes and samples with the help of the ASU/TGen Supercomputer
- Thanks to my collaborators at ASU:
Drs. Zhang, Kim and Baral
- More details available at poster session