Bayesian networks for variable groups

Pekka Parviainen and Samuel Kaski

Aalto University

8.9.2016

Introduction

- Structures of Bayesian networks represent conditional dependencies and independencies between variables
- Question: What can (and cannot) be learned when we replace variables with groups of variables?

Introduction

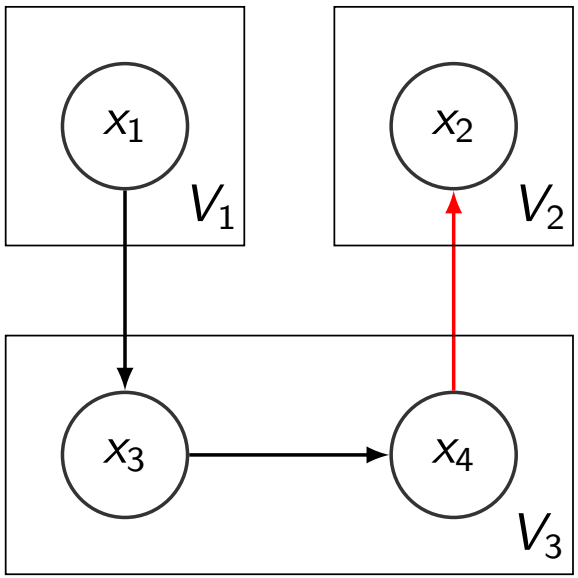
- Structures of Bayesian networks represent conditional dependencies and independencies between variables
- Question: What can (and cannot) be learned when we replace variables with groups of variables?
- Why?
  - Sometimes relations between groups of variables are more interesting than relations between variables
    - For example, multiple different measurements of expression of the same genes, made with multiple measurement platforms
    - Find relationships between the genes and not of the measurement platforms

Preliminaries

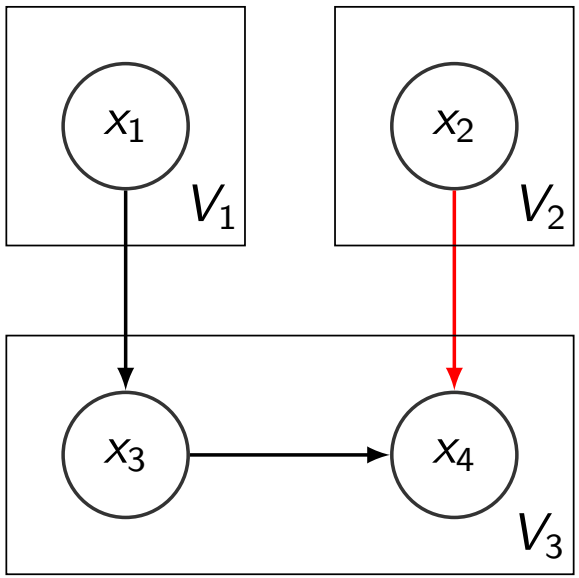
- Terminology:
  - *Variable DAG* = Dependency structure for variables (the "usual" BN structure)
  - *Group DAG* = Dependency structure for variable groups
- Assumptions:
  - Data are generated from a distribution that is faithful to a variable DAG
  - Groups are given
- Methodology: Try to apply standard techniques, see when things break

Groupwise faithfulness

- A distribution p is groupwise faithful to a group DAG H given groups W if it implies exactly the same set of conditional independencies over W as H



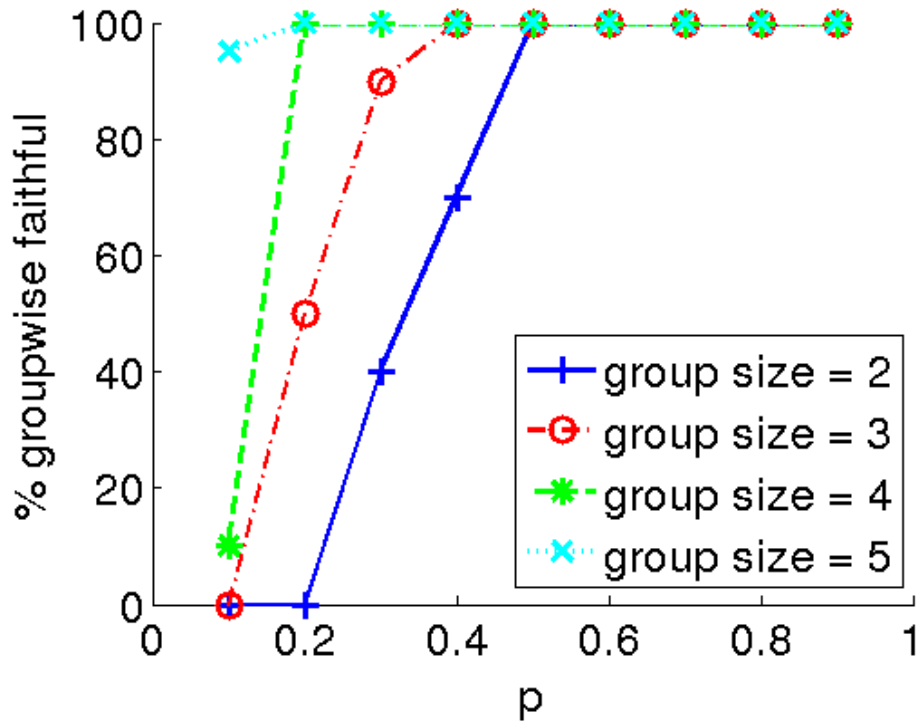
Groupwise faithful



Not groupwise faithful

Strength of the groupwise faithfulness assumption

- Simulation study
- Random DAGs from model $G(20, p)$, groups chosen randomly

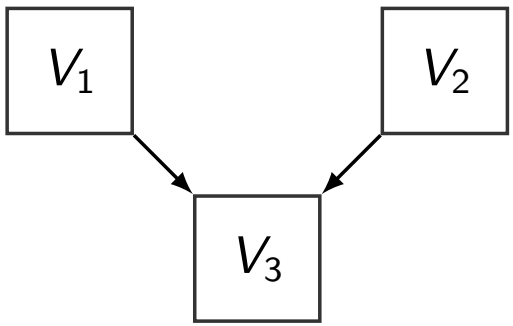


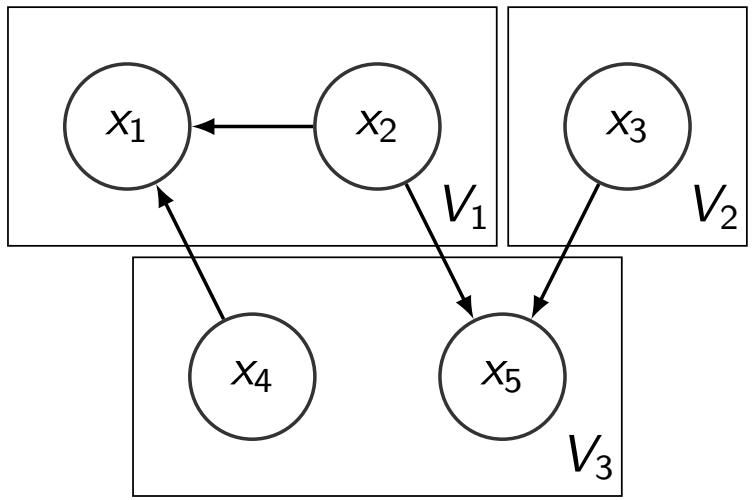
Groupwise causality

- Our definition: Group V is a *group cause* of group U if $P(U|do(V = V_1)) \neq P(U|do(V = V_2))$ for some instantiations V_1 and V_2
- To what extent can group causality be learned using only groupwise independencies (under standard assumptions)?

Group causality (cont.)

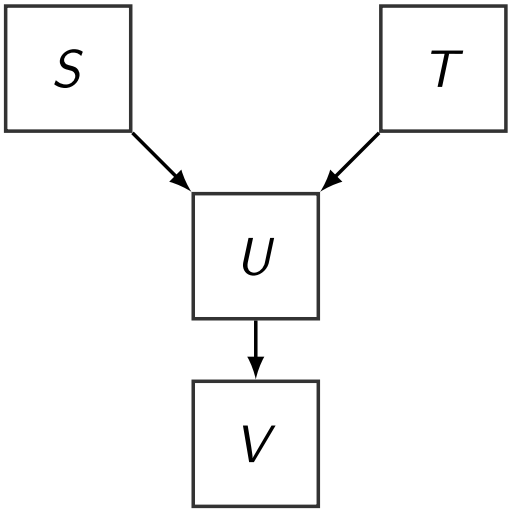
- If there is a v-structure in the group DAG then the corresponding arcs imply group causality

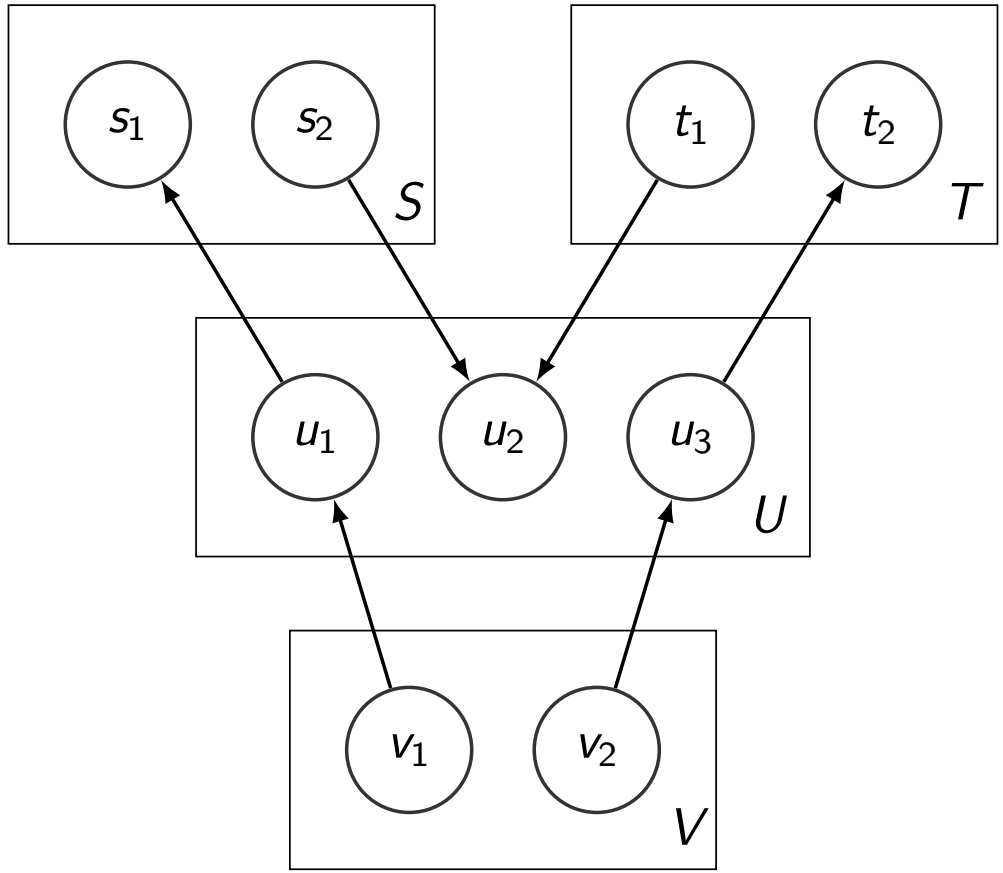




Group causality (cont.)

- Arcs directed according to Meek rules cannot be interpreted group causally





Learning

- Direct learning
  - Create a new variable for each group (Cartesian product)
  - Learn a group DAG from the new variables
- Learning via variable DAGs
  - Learn a variable DAG
  - Infer a group DAG from a variable DAG
- In practice, learning via variable DAGs gives more accurate results

Conclusion

- Group DAGs
  - Express conditional independencies between groups of variables
  - Strong assumptions, theoretical limitations

Thank you!