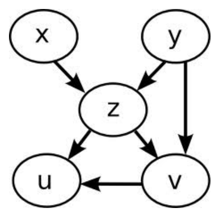
# Bayesian Networks

Guilherme J. M. Rosa

University of Wisconsin-Madison



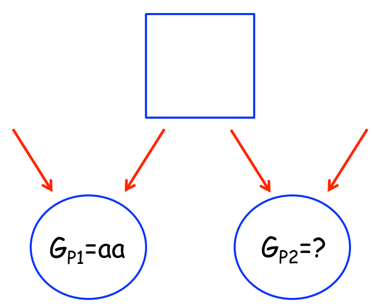

## Marginal and Conditional Independence

$Z \rightarrow X \rightarrow Y$

$Z \leftarrow X \leftarrow Y$

$Z \leftarrow X \rightarrow Y$

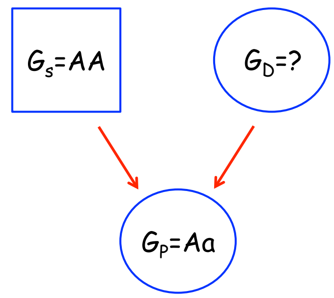
- ✓ Z & Y marginally not independent
- ✓ Conditioned on X they become independent




$Pr(G_{P2}) \neq Pr(G_{P2}|G_{P1})$

Z → X ← Y

- ✓ Z & Y marginally independent
- ✓ Conditioned on X they are not independent
- ✓ Concept of collider, V-structure




$Pr(G_D=aa|G_S=AA) = Pr(G_D=aa) = q^2$

$Pr(G_D=aa|G_S=AA, G_P=Aa) = q \quad [\neq Pr(G_D=aa|G_P=Aa)]$

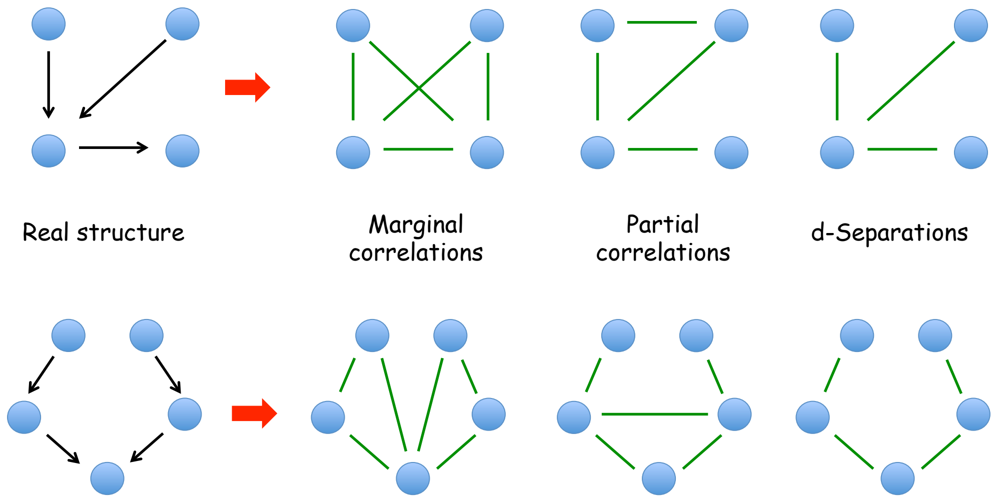
# 'Directed' Separation

⇨ *d-Separation* concept:

Two variables X and Y are said to be d-separated by Q if there is no active path between any X and Y conditionally on Q

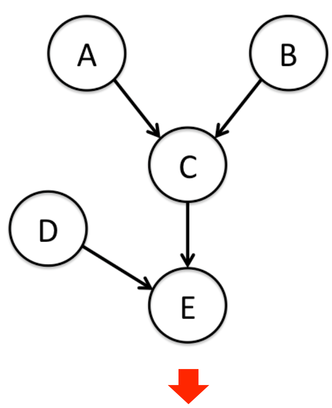
(Verma and Pearl 1988, Pearl 1998, Geiger et al. 1990)

# Examples




# Bayesian Networks

- Graphic representation of a probability distribution over a set of variables → DAG




- Parents: Nodes that are directed to another node(s)
- Child: The descendent node
- Spouse: A node is defined as a spouse when it shares a child with another node

$$\Pr(A,B,C,D,E) = \Pr(E \mid C,D)\Pr(C \mid A,B)\Pr(D)\Pr(B)\Pr(A)$$

## Bayesian Networks

- Usefulness and Applications
  - Flow of information from DNA to phenotype
  - Parsimonious models for multi-trait analysis
  - Prediction, Markov Blanket
  - Causal inference
  - Visualization and model selection tool

## Inference Steps

### ① Structure Learning

- Score-based algorithms
- Constraint-based algorithms

$$\begin{tikzcd} & y_1 \arrow[dr] & \\ y_2 \arrow[rr] & & y_3 \end{tikzcd} \Rightarrow \quad \Lambda = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \lambda_{31} & \lambda_{32} & 0 \end{bmatrix}$$

### ① Parameter Estimation

$$y_3 = \mu + \lambda_{31} y_1 + \lambda_{32} y_2 + e$$

Maximum Likelihood or
Bayesian Inference

# Structure Learning

- Constraint-based algorithms
  - IC, PC – Spirtes et al. (2001)
  - Grow-Shrink (GS) – Margaritis (2003)
  - Incremental Association Markov Blanket (IAMB) – Tsamardinos et al. (2003)
  - Max-Min Parents & Children (MMPC)
- Score-based algorithms
  - Hill Climbing (HC) – Bouckaert (1995)
  - Tabu Search (Tabu)
- Hybrid structure learning algorithms
  - Sparse Candidate (SC) – Friedman et al (1999)
  - Max-Min Hill Climbing (MMHC) – Tsamardinos et al. (2006)

# Constraint-based algorithms

- Series of conditional independence tests (parametric, semiparametric and permutation)
  - Linear correlation or Fisher's Z (continuous data; multivariate normal distribution)
  - Pearson's $X^2$ or mutual information (categorical data; multinomial distribution)
  - Jonckheere-Terpstra (ordinal data)

# Score-based algorithms

- Different score functions
  - Akaike Information Criterion (AIC)
  - Bayesian Information Criterion (BIC)
  - multinomial log-likelihood, Dirichlet posterior density (BDe) or K2 score (categorical data)

# The IC Algorithm
## (Inductive Causation; Verma and Pearl 1991)

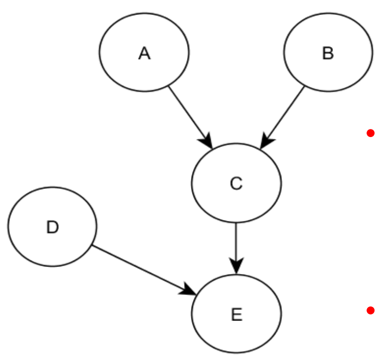
Step 1: Undirected graph (search for d-separations; connect adjacent variables)

Step 2: Partially oriented graph (search for colliders)

Step 3: Attempt to orient remaining undirected edges such that no new colliders or cycles are generated

Step 1: skeleton; Step 2: V structures

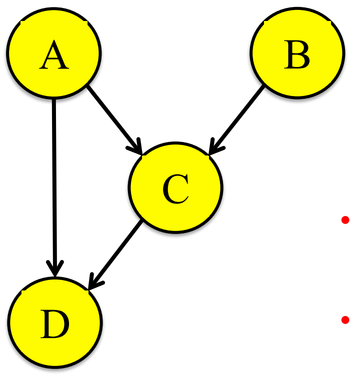
# Markov Blanket




- Markov Blanket (MB): a MB of a node is defined as the set containing its parent(s), child(ren) and spouse(s)
- Conditionally on its MB, a node is independent from all other nodes

Examples: MB(D) = {C,E}; MB(E) = {C,D}; MB(C) = {A,B,D,E}

## Causal Inference




- Arrows: Causal interpretation; consequences of intervention
- Direct, indirect and total effects
- Additional assumptions: Markov condition, faithfullness and causal sufficiency assumptions

## Causal Inference

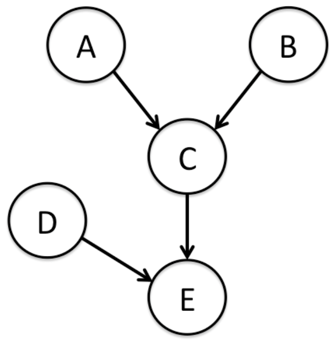
- Prediction of the result of an intervention (gene knockout, management decision, treatment effect)
- Estimation of causal effects:

  If the causal DAG is known and the distribution is multivariate Gaussian, then the causal effect (β) of X on Y can be estimated from the regression :

  $$E[Y] = m + \beta X + pa(X)$$

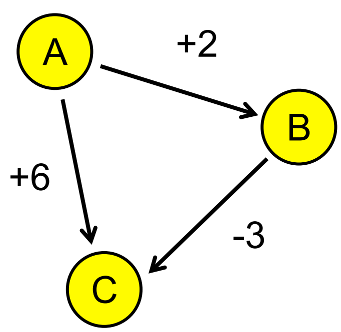
- i.e., DAG determines adjustment variables [backdoor adjustment; Pearl (1993)]

## Markov Condition




$$\Pr(A, B, C, D, E) = \Pr(E \mid C, D)\Pr(C \mid A, B)\Pr(D)\Pr(B)\Pr(A)$$

## Faithfullness




## Causal Sufficiency