Graphical Models

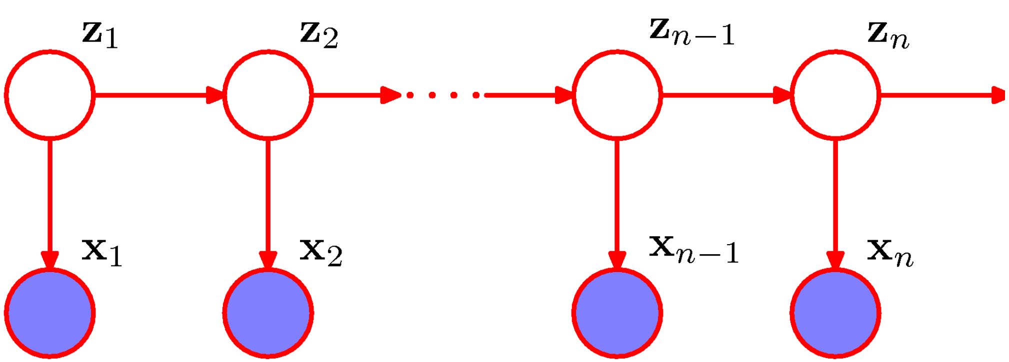
Aarti Singh
Slides Courtesy: Carlos Guestrin

Machine Learning 10-701/15-781
Apr 12, 2023

iid to dependent data

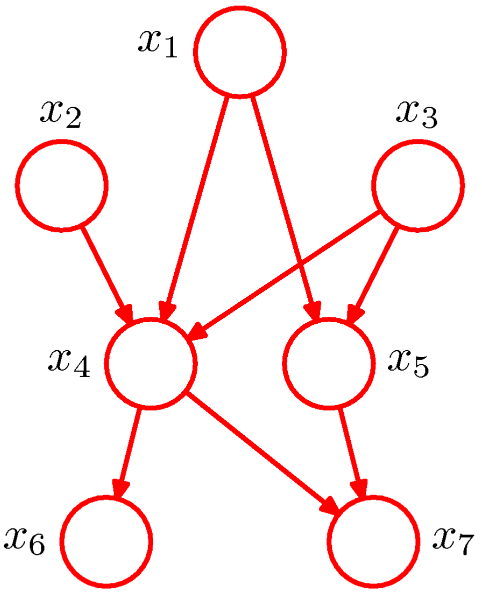
HMM

- sequential dependence



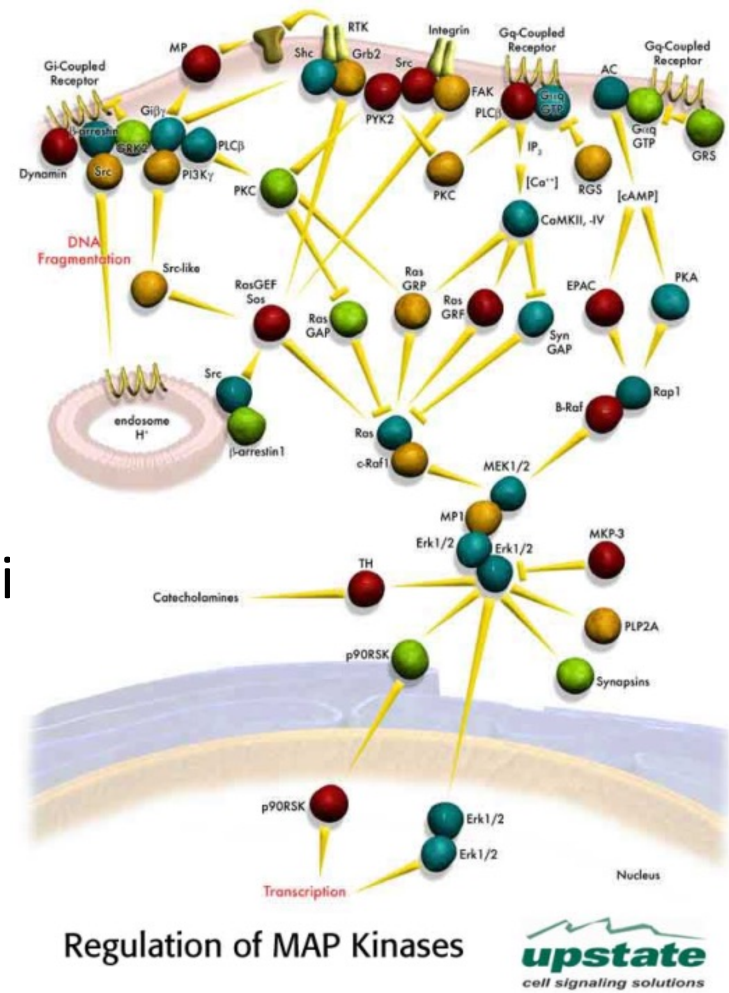
Graphical Models

- general conditional dependence



Applications

- Diagnosis of diseases
- Study Human genome
- Robot mapping
- Brain networks
- Fault diagnosis
- Modeling sensor network data
- Modeling protein-protein interacti
- Weather prediction
- Computer vision
- Statistical physics
- Many, many more …



Conditional Independence

- X is **conditionally independent** of Y given Z:

  probability distribution governing X is independent of the value of Y, given the value of Z

$$(\forall x, y, z) P(X = x | Y = y, Z = z) = P(X = x | Z = z)$$

- Equivalent to:

$$P(X, Y \mid Z) = P(X \mid Z) P(Y \mid Z)$$

- Also to:

$$P(X \mid Y, Z) = P(X \mid Z)$$

Graphical Models

- Key Idea:
  - Conditional independence assumptions useful
  - but Naïve Bayes is extreme!
  - Graphical models express sets of conditional independence assumptions via graph structure
  - **Graph structure** + **Conditional Probability Tables (CPTs)** define *joint probability distribution over set of variables/nodes*

- Two types of graphical models:
  - Directed graphs (aka Bayesian Networks) Today
  - Undirected graphs (aka Markov Random Fields)

Topics in Graphical Models

- Representation
  - Which joint probability distributions does a graphical model represent?
- Inference
  - How to answer questions about the joint probability distribution?
    - Marginal distribution of a node variable
    - Most likely assignment of node variables
- Learning
  - How to learn the parameters and structure of a graphical model?

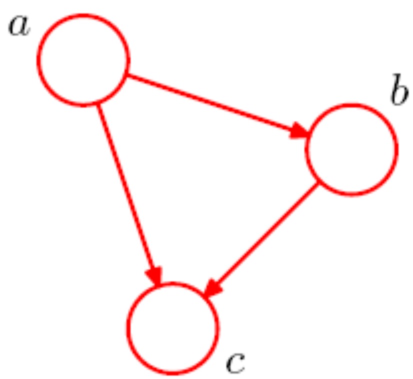
Directed - Bayesian Networks

- Representation
  - Which joint probability distributions does a graphical model represent?

For any arbitrary distribution,

Chain rule:

$$p(a, b, c) = p(c|a, b)p(b|a)p(a)$$



More generally:

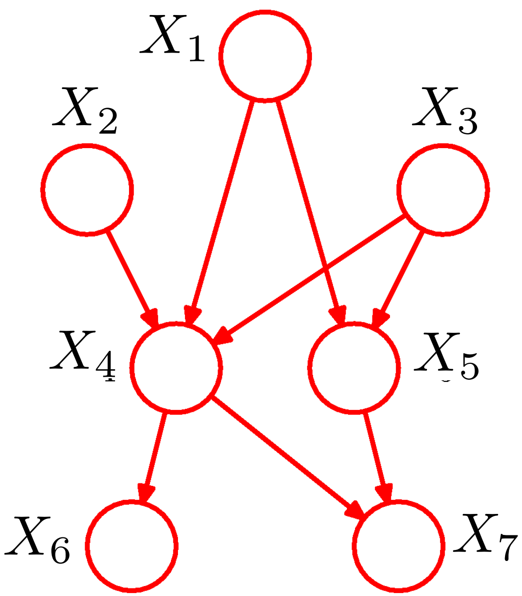
$$p(\mathbf{X}) \quad = \quad \prod_{i=1}^{n} p(X_i | X_{i-1}, \ldots, X_1)$$

Fully connected directed graph between $X_1, \ldots, X_n$

Directed - Bayesian Networks

- Representation
  - Which joint probability distributions does a graphical model represent?

Absence of edges in a graphical model conveys useful information.



$$p(X_1, \ldots, X_7) =$$
$$p(X_1)p(X_2)p(X_3)p(X_4|X_1, X_2, X_3) \cdot$$
$$p(X_5|X_1, X_3)p(X_6|X_4)p(X_7|X_4, X_5)$$

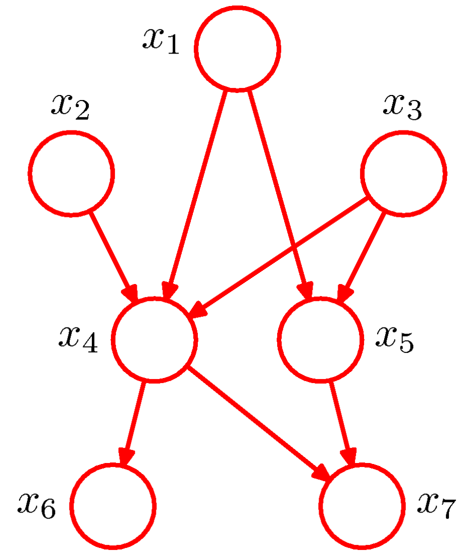
Directed – Bayesian Networks

- Compact representation for a joint probability distribution
- Bayes Net = Directed Acyclic Graph (DAG) + Conditional Probability Tables (CPTs)
- distribution factorizes according to graph

$$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$

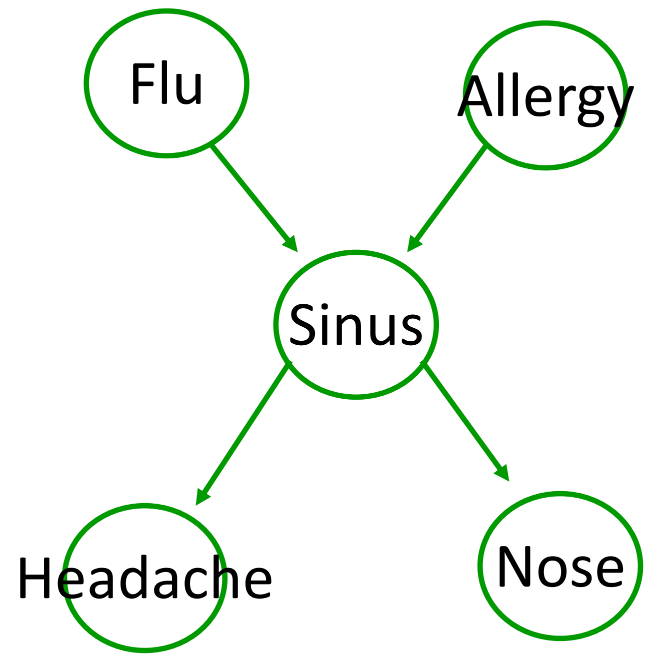
≡ distribution satisfies **local Markov assumption**

x_k is independent of its non-descendants given its parents pa_k



Bayesian Networks Example

- Suppose we know the following:
  - The flu causes sinus inflammation
  - Allergies cause sinus inflammation
  - Sinus inflammation causes a runny nose
  - Sinus inflammation causes headaches



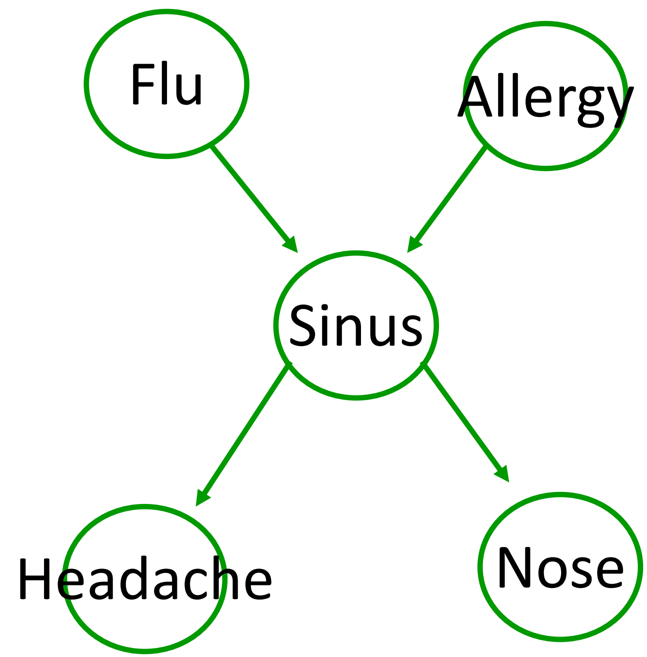
- Causal Network

- Local Markov Assumption: If you have no sinus infection, then flu has no influence on headache (flu causes headache but only through sinus)

Markov independence assumption

Local Markov Assumption: A variable X is independent of its non-descendants given its parents (only the parents)

	parents	non-desc	assumption
S	F,A	-	-
H	S	F,A,N	$H \perp \{F,A,N\} \mid S$
N	S	F,A,H	$N \perp \{F,A,H\} \mid S$
F	-	A	$F \perp A$
A	-	F	$A \perp F$



Markov independence assumption

Local Markov Assumption: A variable X is independent of its non-descendants given its parents (only the parents)

Joint distribution:

P(F, A, S, H, N)

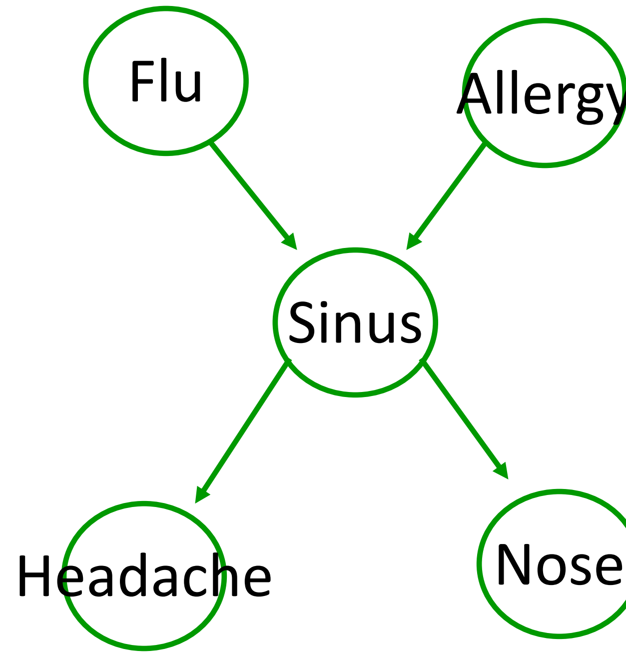
= P(F) P(A|F) P(S|F,A) P(H|S,F,A) P(N|S,F,A,H)

Chain rule

= P(F) P(A) P(S|F,A) P(H|S) P(N|S)

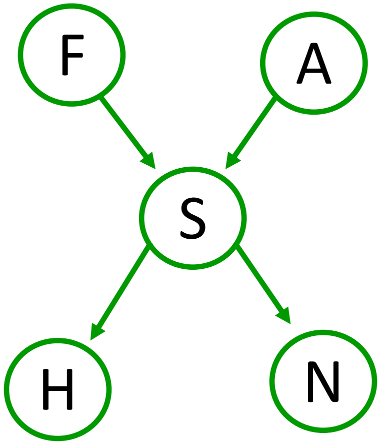
Markov Assumption

$F \perp A, \quad H \perp \{F,A\}|S, \quad N \perp \{F,A,H\}|S$



Bayesian Network - ingredients

- Discrete variables $X_1, \ldots, X_n$
- Directed Acyclic Graph (DAG)
  - Defines parents of X_i, $\mathbf{Pa}_{X_i}$
- CPTs (Conditional Probability Tables)
  - $P(X_i \mid \mathbf{Pa}_{X_i})$



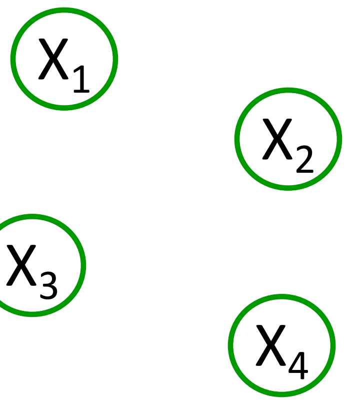
E.g. $X_i = S$, $\mathbf{Pa}_{X_i} = \{F, A\}$

	F=f, A=f	F=t, A=f	F=f, A=t	F=t,A=t
S=t	0.9	0.8	0.7	0.3
S=f	0.1	0.2	0.3	0.7

n variables, K values, max d parents/node $O(nK \times K^d)$

Two (trivial) special cases

Fully disconnected graph



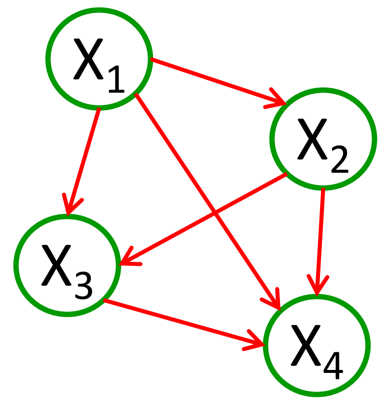
X_i

parents: ϕ

non-descendants: $X_1,\ldots,X_{i-1}$, $X_{i+1},\ldots, X_n$

$X_i \perp X_1,\ldots,X_{i-1},X_{i+1},\ldots, X_n$

Fully connected graph



X_i

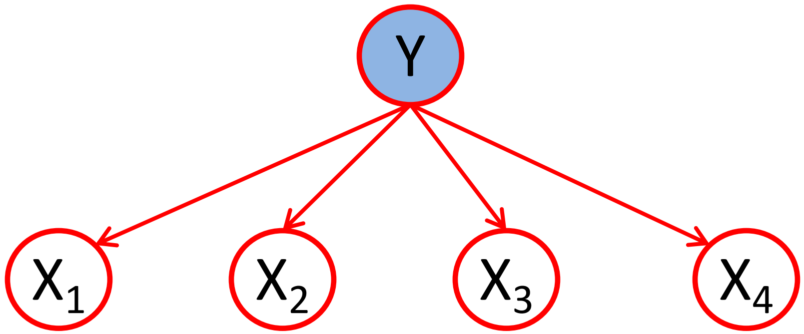
parents: $X_1, \ldots, X_{i-1}$

non-descendants: ϕ

No independence assumption

Bayesian Networks Example

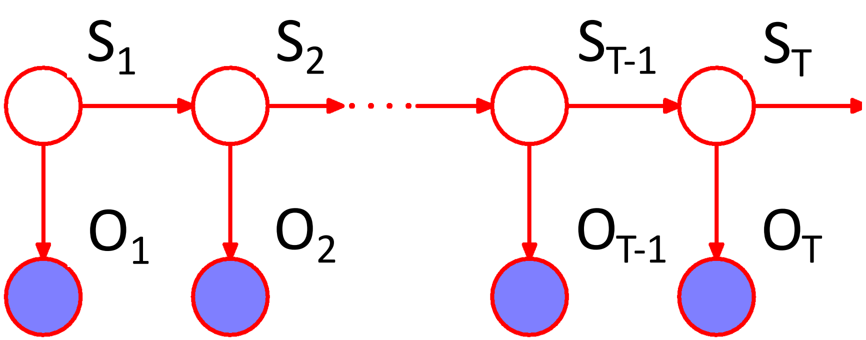
- Naïve Bayes $X_i \perp X_1,\ldots,X_{i-1},X_{i+1},\ldots, X_n | Y$



$P(X_1,\ldots,X_n,Y) =$

$P(Y)P(X_1|Y)\ldots P(X_1|Y)$

- HMM



$$p(\{S_t\}_{t=1}^T, \{O_t\}_{t=1}^T) \quad = \quad p(S_1)\prod_{t=2}^T p(S_t|S_{t-1})\prod_{t=1}^T p(O_t|S_t)$$

Explaining Away

Local Markov Assumption: A variable X is independent of its non-descendants given its parents (only the parents)

$F \perp A$ $\quad\quad$ $P(F|A=t) = P(F)$

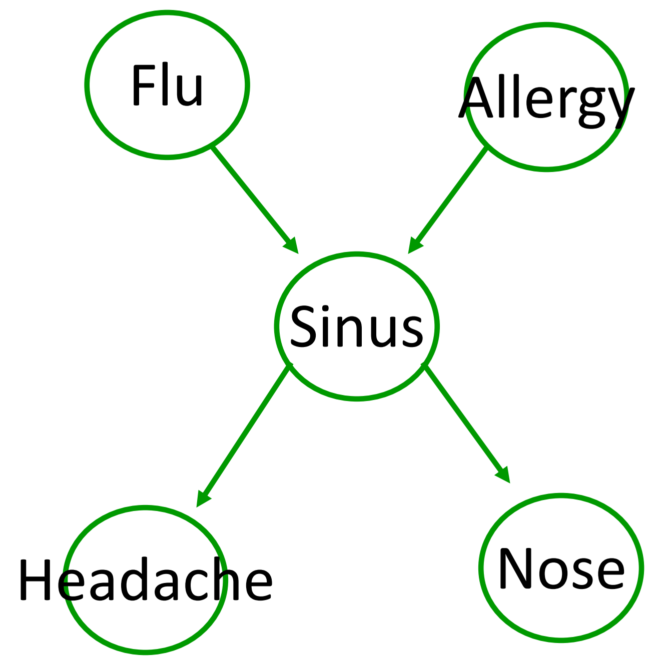
$F \perp A|S$?
$P(F|A=t,S=t) = P(F|S=t)$?

No!

$P(F=t|S=t)$ is high,
but $P(F=t|A=t,S=t)$ not as high
since $A = t$ explains away $S=t$

Infact, $P(F=t|A=t,S=t) < P(F=t|S=t)$

$F \perp A|N$? $\quad$ No!



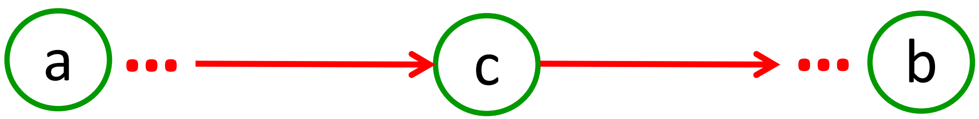
Independencies encoded in BN

- We said: All you need is the local Markov assumption
  - $(X_i \perp \text{NonDescendants}_{Xi} \mid \mathbf{Pa}_{Xi})$
- But then we talked about other (in)dependencies
  - e.g., explaining away

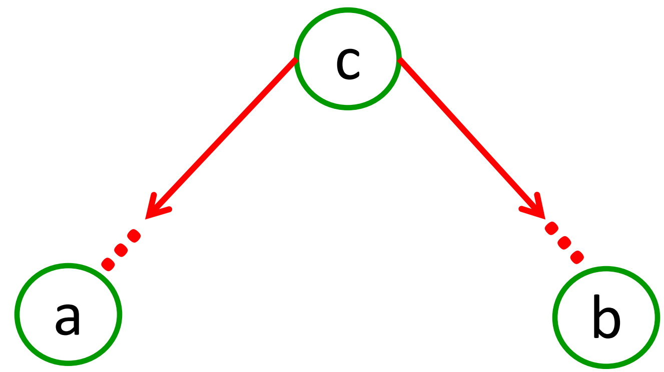
- What are the independencies encoded by a BN?
  - Only assumption is local Markov
  - But many others can be derived using the algebra of conditional independencies!!!

D-separation

- a is D-separated from b by c ≡ $a \perp b | c$
- Three important configurations

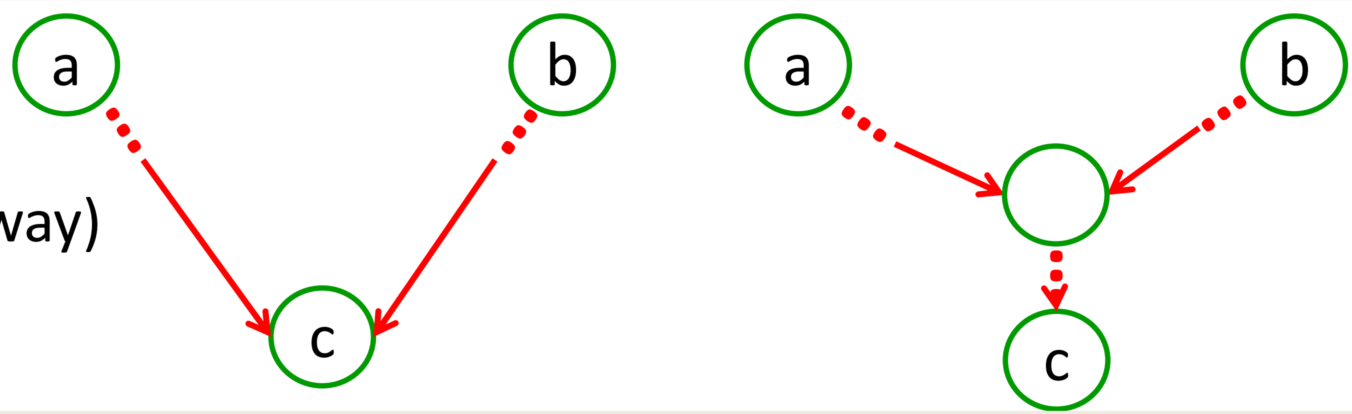


Causal direction $a \perp b | c$



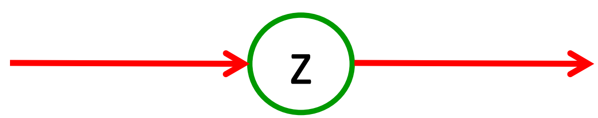
Common cause $a \perp b | c$

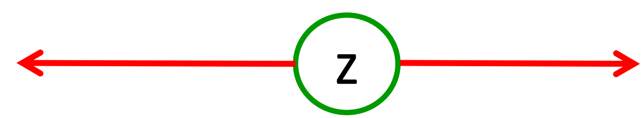
V-structure
(Explaining away)
$a \not\perp b | c$



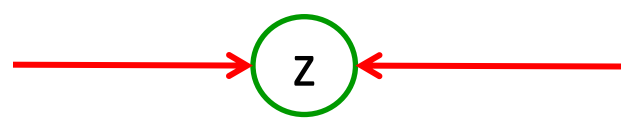
D-separation

- A, B, C – non-intersecting set of nodes
- A is D-separated from B by C $\equiv A \perp B | C$
  if all paths between nodes in A & B are “blocked”
  i.e. path contains a node z such that either





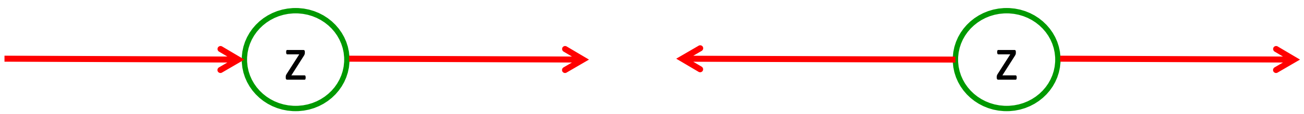
and z in C, OR



and neither z nor any of its descendants is in C.

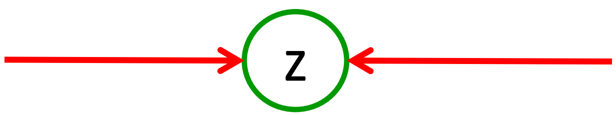
D-separation Example

A is D-separated from B by C if every path between A and B contains a node z such that either

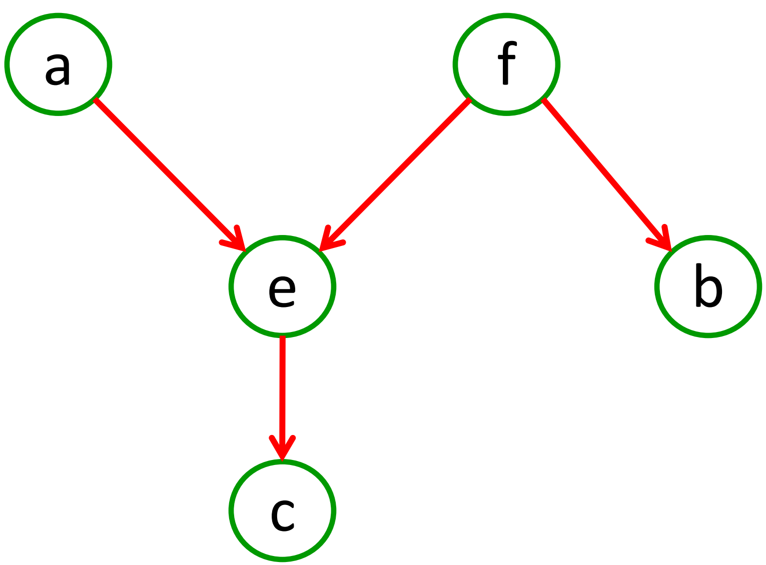


And z in C

or



And neither z nor its descendants are in C



$a \perp b \mid f$?
Yes, Consider z = f or z = e

$a \perp b \mid c$?
No, Consider z = e

- Poll

Representation Theorem

- Set of distributions that factorize according to the graph - F

- Set of distributions that respect conditional independencies implied by d-separation properties of graph – I

$$I \Rightarrow F$$

Important because: **Given independencies of *P* can get BN structure *G***

$$I \Leftarrow F$$

Important because: **Read independencies of *P* from BN structure *G***

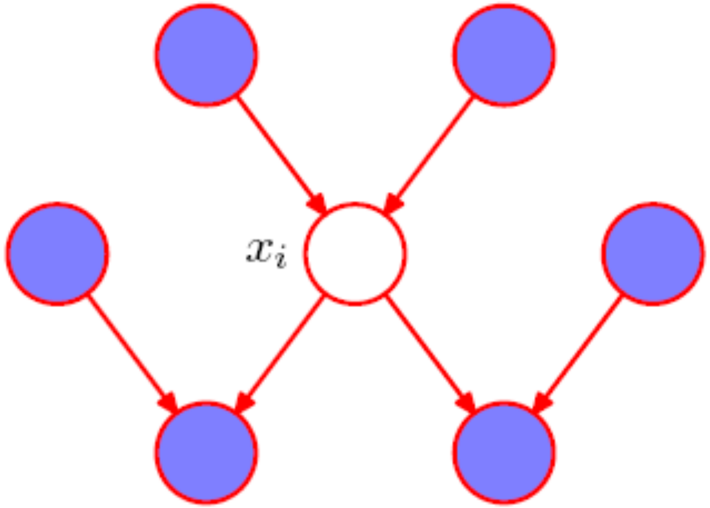
Markov Blanket

- Conditioning on the Markov Blanket, node i is independent of all other nodes.

$$p(\mathbf{x}_i|\mathbf{x}_{\{j\neq i\}}) = \frac{p(x_1,\ldots,x_n)}{\sum_i p(x_1,\ldots,x_n)} = \frac{\prod_k p(x_k|pa(x_k))}{\sum_i \prod_k p(x_k|pa(x_k))} = p(\mathbf{x}_i|\mathrm{MB}(\mathbf{x}_i))$$

Only terms that remain are the ones which involve i

$$p(x_i|pa(x_i)) \qquad p(x_k|pa(x_k) \ni i)$$



- Markov Blanket of node i - Set of parents, children and co-parents of node i

Directed – Bayesian Networks

- Graph encodes local independence assumptions (local Markov Assumptions)
- Other independence assumptions can be read off the graph using d-separation
- distribution factorizes according to graph ≡ distribution satisfies all independence assumptions found by d-separation

$$F \Longleftrightarrow I$$

- Does the graph capture all independencies? Yes, for *almost all* distributions that factorize according to graph. More in 10-708