Graphical Models

Lecture 2:

Bayesian Network Representation

Andrew McCallum
mccallum@cs.umass.edu

Thanks to Noah Smith and Carlos Guestrin for some slide materials.

Administrivia

- This course “likely” but not “certain”
  to be an AI core.
  Won’t know for sure until February 2nd.

- Mailing list 691gm-staff@cs.umass.edu now exists.
  Later 691gm-all@cs will work also.

- Who has visited the web site?
  http://www.cs.umass.edu/~mccallum/courses/gm2011

Goals for Today

- Define Bayesian Networks
- Naive Bayes
- Relation between BNs and independence
- V-structure, active trail, D-separation, Bayes ball
- I-Map, Minimal I-Map, P-Map.

- HW#1 out.

The Bayesian Network Independence Assumption

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

$$X \perp \mathbf{NonDescendants}(X) \mid \mathbf{Parents}(X)$$

- P "factorizes over graph G" defined by **Parents**()

Recipe for a Bayesian Network

- Set of random variables **X**
- Directed acyclic graph (each X_i is a vertex)
- Conditional probability tables, P(X | **Parents**(X))
- Joint distribution:

$$P(\boldsymbol{X}) = \prod_{i=1}^{n} P(X_i \mid \mathbf{Parents}(X_i))$$

- Local Markov Assumption
  - A variable X is independent of its non-descendants given its parents (and *only* its parents).

$$X \perp \mathbf{NonDescendants}(X) \mid \mathbf{Parents}(X)$$

Draw!

Talk about "generative storyline"

Where do Independencies Come From?

- Derive complete set from true P.
  - Generally impossible.
- Brazen convenience.
- Intuition about causality.
- Careful search
  - Structure Learning (later in the semester)

Naive Bayes

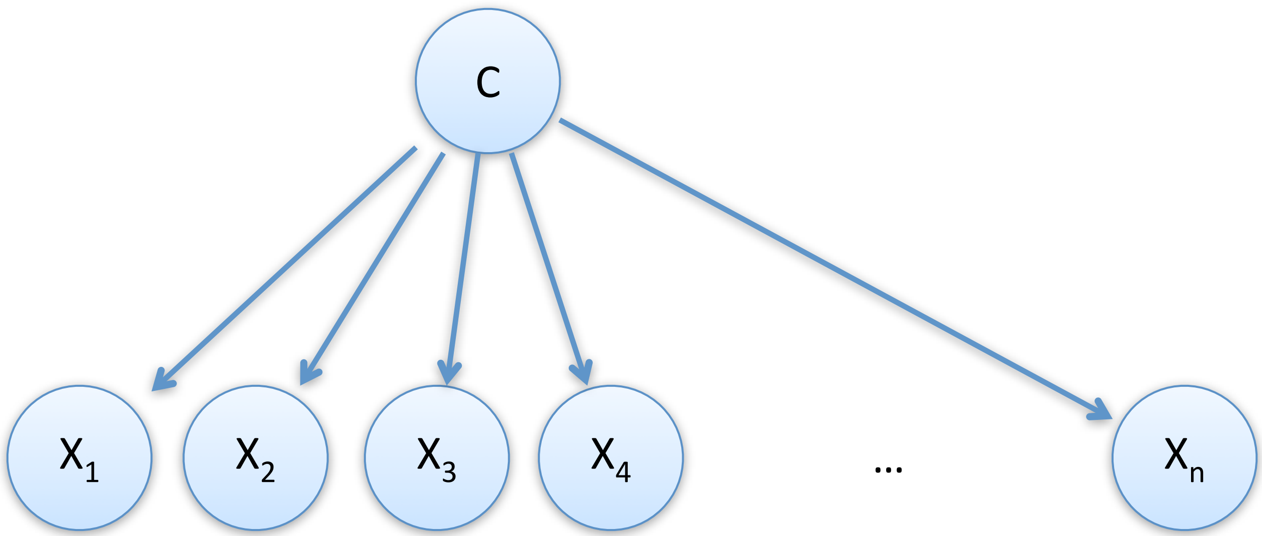
Common, simple independence assumption

Naïve Bayes Model

- Class variable C
- Evidence variables **X** = $X_1, X_2, \ldots, X_n$
- Assumption: $(X_i \perp X_j \mid C)\ \forall\, X_i \subseteq \mathbf{X},\ X_{j \neq i} \subseteq \mathbf{X}$

$$P(C, \boldsymbol{X}) = P(C) \prod_{i=1}^{n} P(X_i \mid C)$$

Naïve Bayes Model



Where do Independencies Come From?

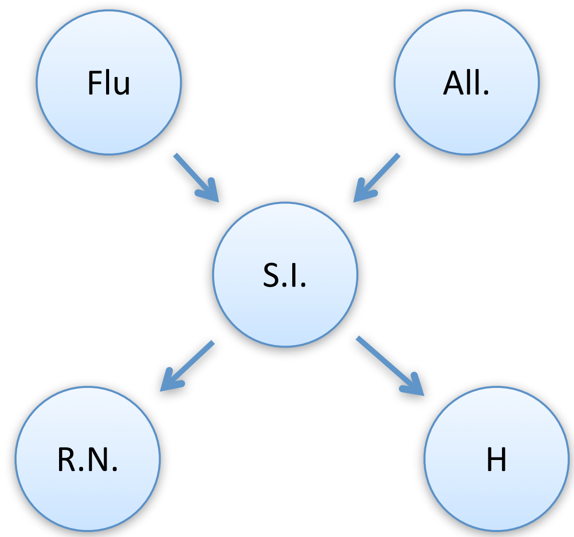
- Derive complete set from true P.
  - Generally impossible.
- Brazen convenience.
- Intuition about causality.
- Careful search
  - Structure Learning (later in the semester)

Causal Structure

- The flu causes sinus inflammation
- Allergies *also* cause sinus inflammation
- Sinus inflammation causes a runny nose
- Sinus inflammation causes headaches

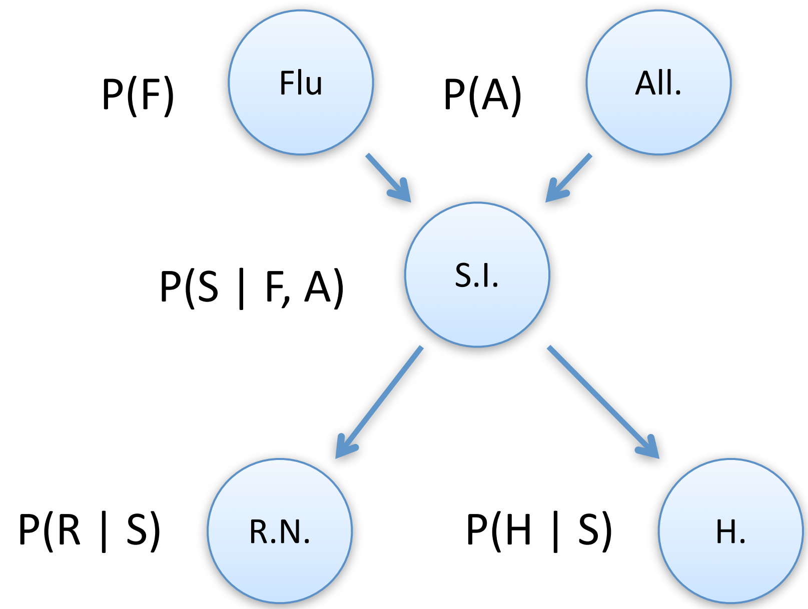
Causal Structure

- The flu causes sinus inflammation
- Allergies *also* cause sinus inflammation
- Sinus inflammation causes a runny nose
- Sinus inflammation causes headaches

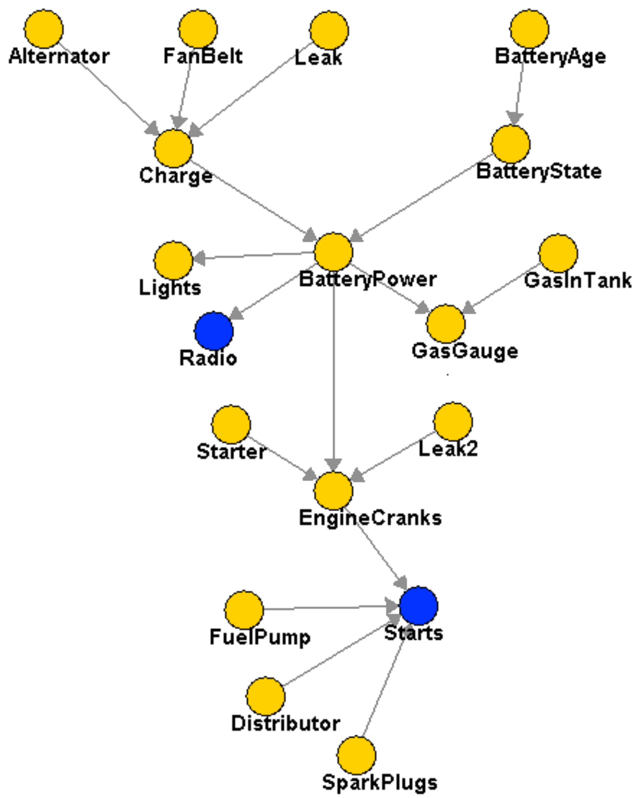


Factored Joint Distribution

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



A Bigger Example: Starting Car



- 18 variables
- The car doesn't start. The radio works.
- What do we conclude about the "gas in tank"?

Causality and Independence

- "A causes B" *implies*
  "A and B dependent"
- "A and B dependent" *does not imply*
  "A causes B"

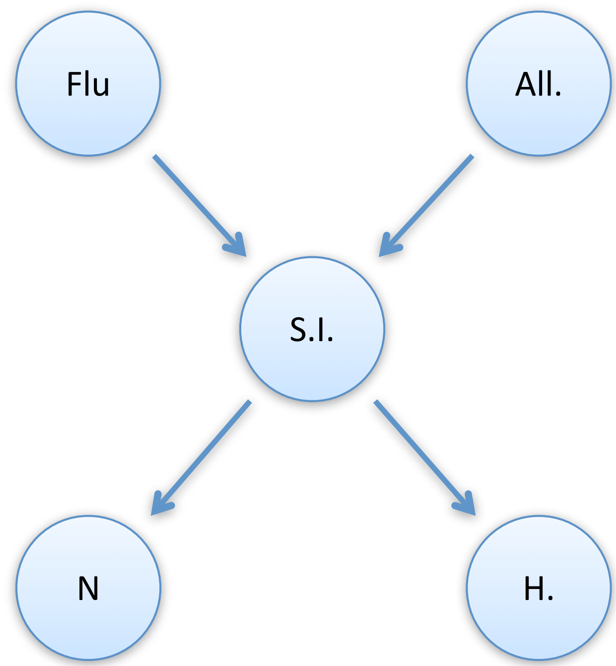
e.g. smoking, cancer, yellow-fingers

Querying the Model

- **Marginal Inference**
  P(F) or P(F|H=t)

- **MAP* Inference**
  argmax $_{f,a}$ P(F=f, A=a|H=t)

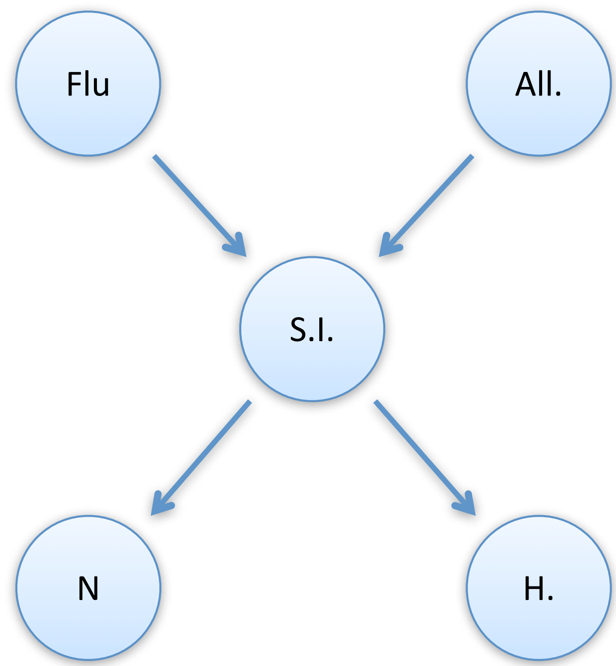
- **Active data collection**
  In solving one of the two above problems, which variable to query next.



* "Maximum Aposteriori," also sometimes called "MPE Inference" (Most Probable Explanation)

Queries and Reasoning Patterns

- **Causal Reasoning** or **Prediction** (downstream)
- **Evidential Reasoning** (upstream)
- **Inter-causal Reasoning** (sideways between parents) "explaining away"

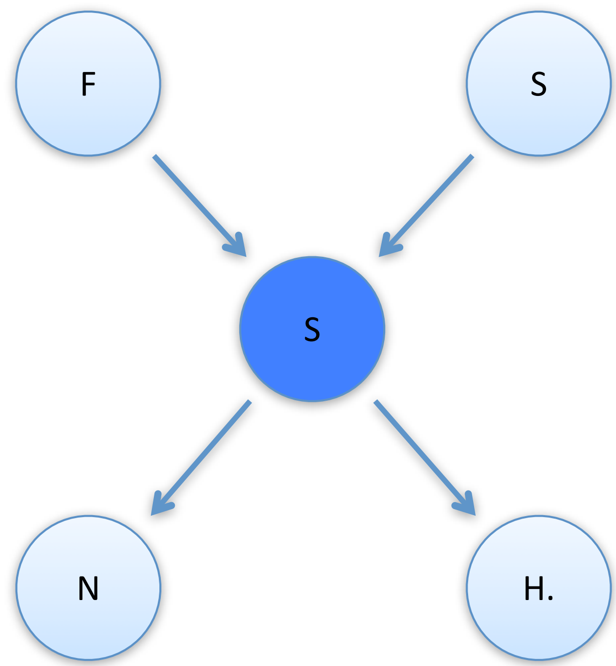


Nothing magical. Underneath everything comes from joint *P* table.

Reading Independencies from the Graph

We used some independencies when building the BN.
Once built the BN expresses some independencies itself. How do we read these from the graph?

- $N \perp F \mid S$
  (follows from BN def'n)



Can we judge independence by the existence of paths with no "blocking" observed variables?

The BN Independence Assumption

again

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

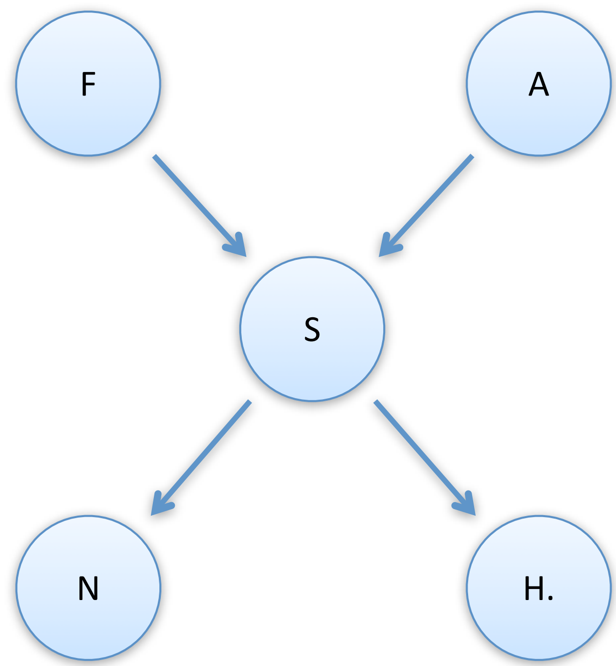
$$X \perp \textbf{NonDescendants}(X) \mid \textbf{Parents}(X)$$

Reading Independencies from the Graph

We used some independencies when building the BN.
Once built the BN expresses some independencies itself. How do we read these from the graph?

- $R \perp F \mid S$
  (follows from BN def'n)

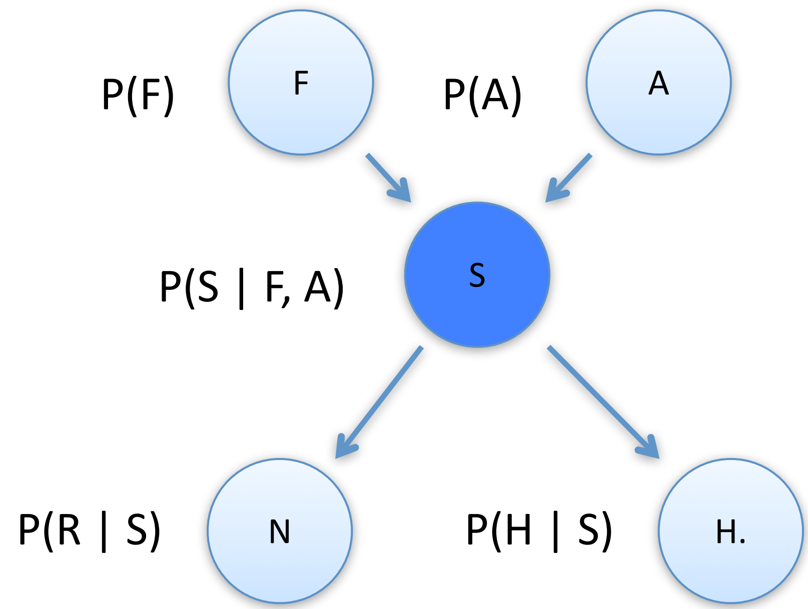
- $R \perp F \mid \varnothing$?
- Answer
  - Can we imagine a case in which independence does not hold? (reason by converse)



Can we judge independence by the existence of paths with no "blocking" observed variables?

A Puzzle

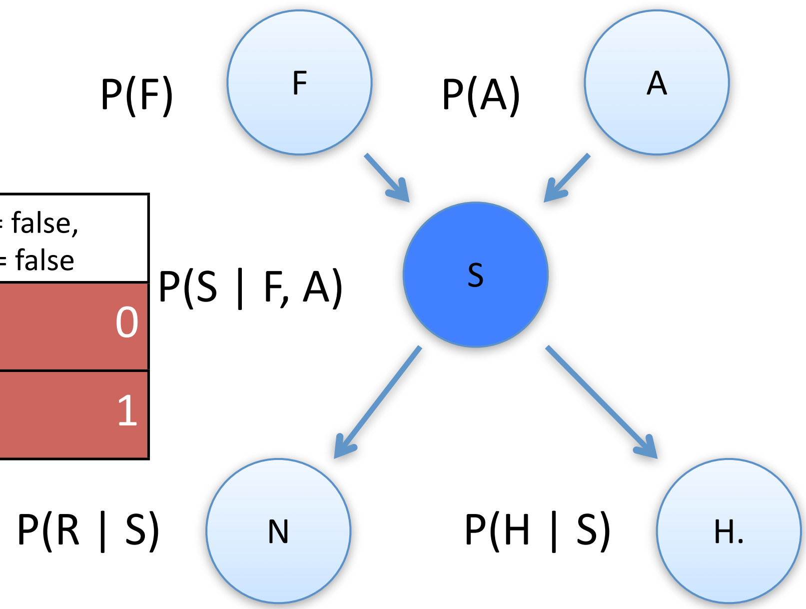
- $F \perp A \mid S$?



A Puzzle

- F ⊥ A | S ?

P(S | F, A)	F = true, A = true	F = true, A = false	F = false, A = true	F = false, A = false
S = true	0	1	1	0
S = false	1	0	0	1



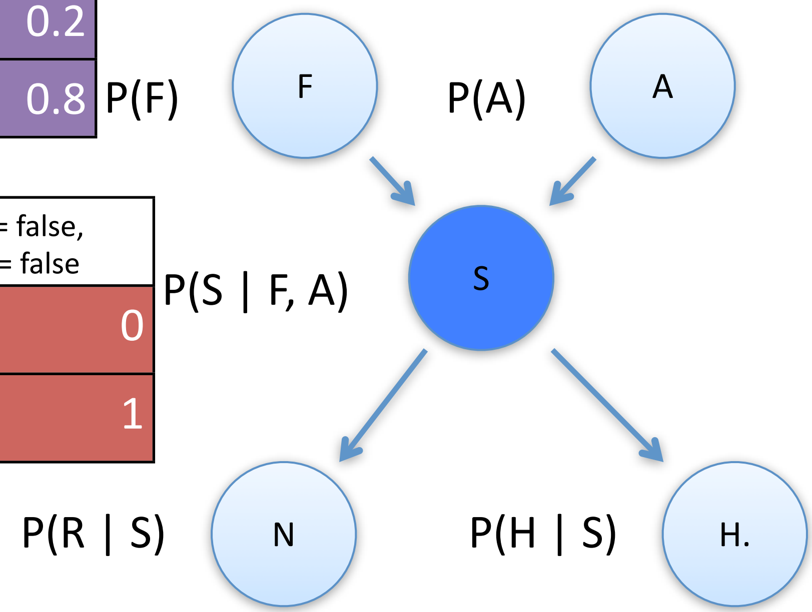
A Puzzle

true	0.2
false	0.8

true	0.2
false	0.8

- F ⊥ A | S ?

P(S|F, A)	F = true, A = true	F = true, A = false	F = false, A = true	F = false, A = false
S = true	0	1	1	0
S = false	1	0	0	1

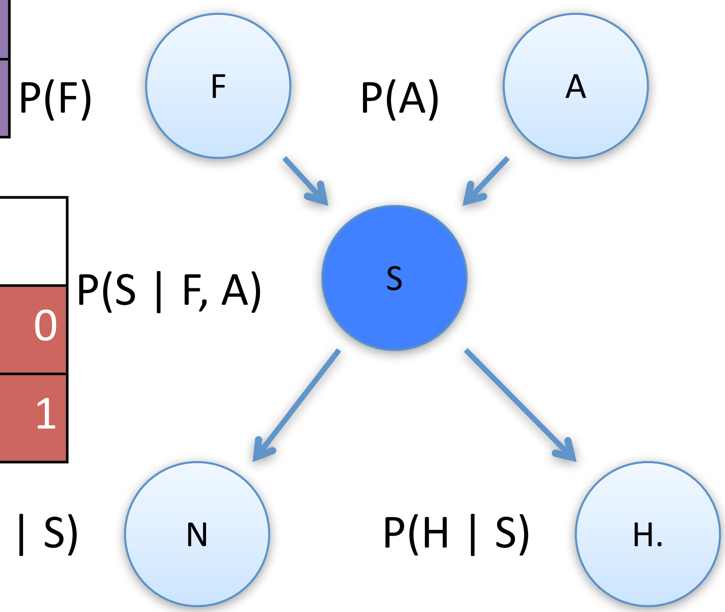


A Puzzle

true	0.2
false	0.8

true	0.2
false	0.8

P(F)



- F ⊥ A | S ?

P(S|F, A)	F = true, A = true	F = true, A = false	F = false, A = true	F = false, A = false
S = true	0	1	1	0
S = false	1	0	0	1

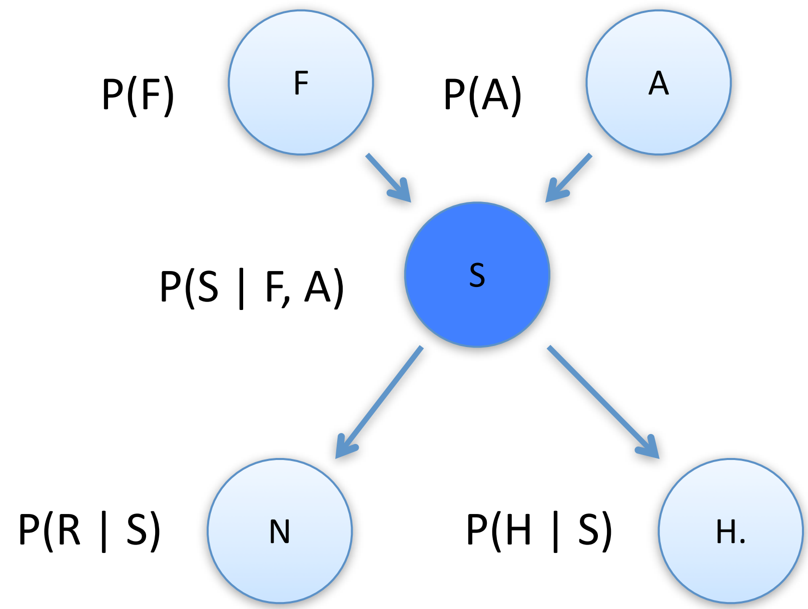
- P(F = true) = 0.2
- P(F = true | S = true) = 0.5
- P(F = true | S = true, A = true) = 0

A Puzzle

- $F \perp A \mid S$?

- In general, **no**.
  - This independence statement does not follow from the Local Markov assumption.

- $\neg (F \perp A \mid S)$



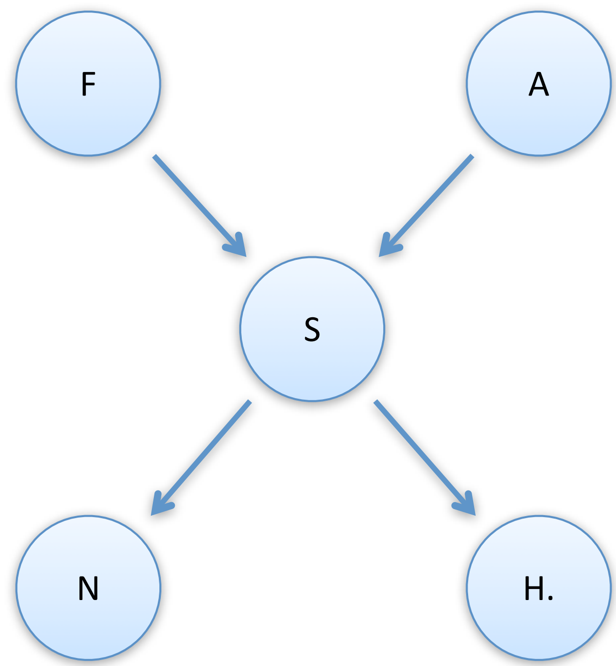
Discuss Pearl's "Alarm" network
"Explaining away"

Reading Dependencies from the Graph

We used some independencies when building the BN.
Once built the BN expresses some independencies itself. How do we read these from the graph?

- **Direct Connection**
  $F \rightarrow S$, $\neg F \perp S$
- **Indirect Causal Effect**
  $F \rightarrow S \rightarrow H$, $\neg F \perp H$
- **Indirect Evidential Effect**
  $H \leftarrow S \leftarrow F$, $\neg H \perp F$
- **Common Cause**
  $N \leftarrow S \rightarrow H$, $\neg N \perp H$
- **Common Effect**
  $F \rightarrow S \leftarrow A$, *S observed* $\neg F \perp A \mid S$

S unobserved



Reading Independencies from the Graph

¬ F ⊥ S | **F ⊥ S**

Direct Connection

X → Y

Indirect Causal Effect

X → Z → Y

Indirect Evidential Effect

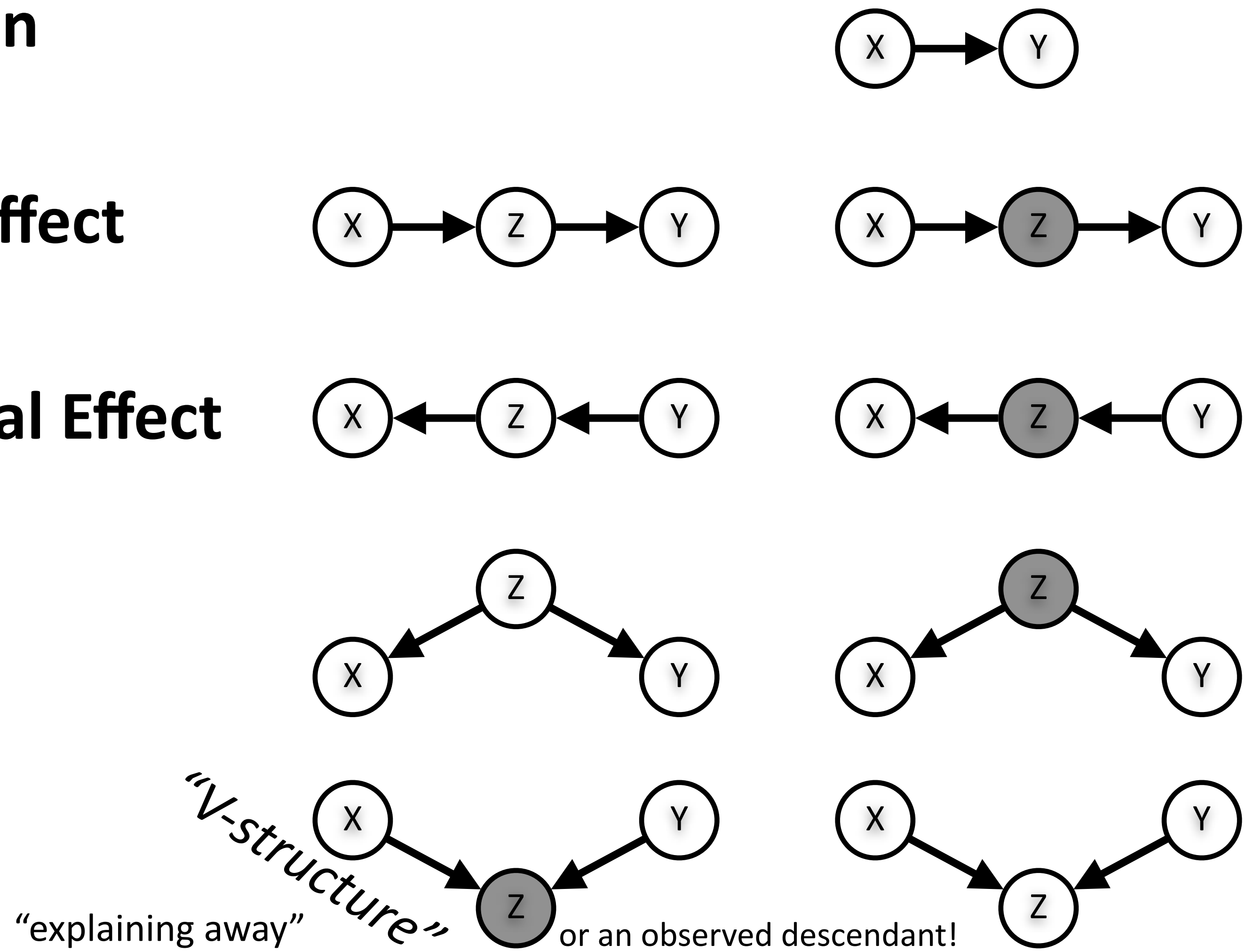
X ← Z ← Y

Common Cause

X ← Z → Y

Common Effect

X → Z ← Y



“Active trail” ~ dependent

Active Trail

Let *G* be a BN structure and $X_1 \Leftrightarrow \ldots \Leftrightarrow X_n$ a trail in *G*.

Let ***Z*** be a subset of observed variables.

The trail is “***active***” given ***Z*** if

- whenever we have a *v-structure* $X_{i-1} \rightarrow X_i \rightarrow X_{i+1}$, then X_i or one of its descendants are in ***Z***;
- no other node along the trail is in ***Z***.

"Separated" ~ independent

D-Separation

"Directed Separation"

Let **X**, **Y**, **Z** be three sets of nodes in *G*. We say that **X** and **Y** are *"d-separated"* given **Z**,

$$\text{d-sep}_G(\mathbf{X} ; \mathbf{Y} \mid \mathbf{Z}),$$

if there is no *"active trail"* between any node X ∈ **X** and Y ∈ **Y** given **Z**.

D-Separation Algorithm

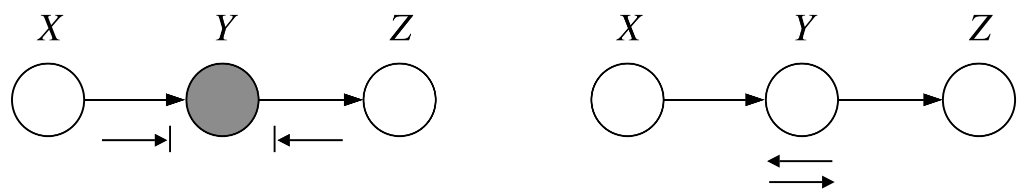
- Question: Are X and Y d-separated given **Z**?
  – (How many possible trails?)

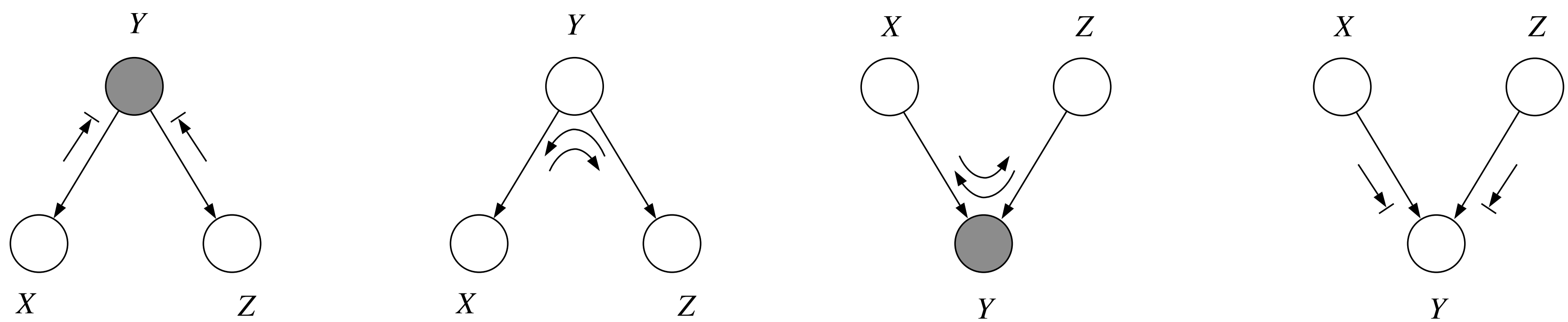
1. Traverse the graph bottom up, marking any node that is in **Z** or with a descendent in **Z**.
2. Breadth-first search from X, only along active trails; finds reachable set **R**.
   – Extra bookkeeping required to keep track of each node being reached via children vs. via parents!
3. X and Y are d-separated iff Y $\notin$ **R**.

Bayes Ball Algorithm (due to Ross Shachter)

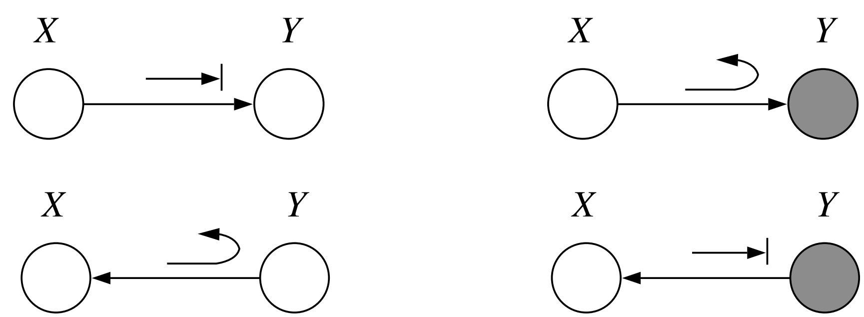
Another expression of “active trails” and d-separation.

Behavior going from X to Z on path through Y.





Behavior at end points.



Diagrams from M. Jordan

D-Separation and Dependencies

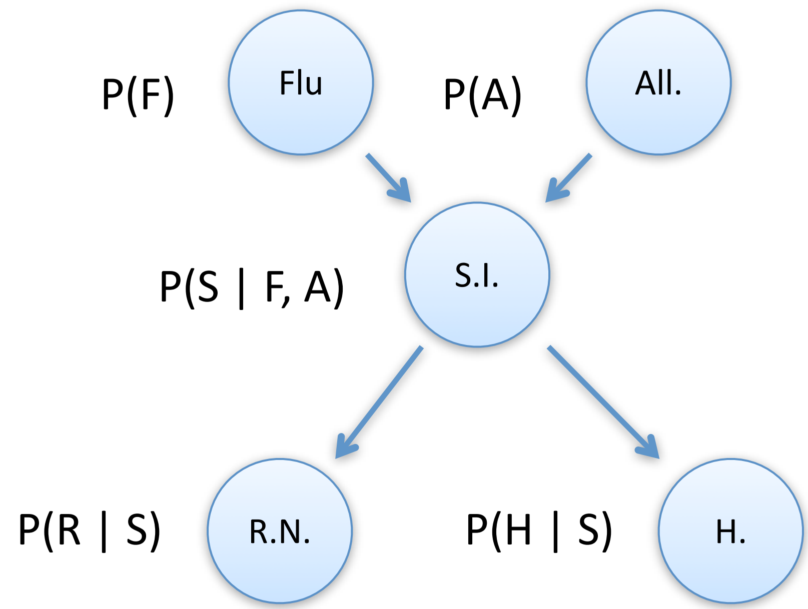
Theorem 3.4, (K&F p73):
Let *G* be a BN structure. If *X* and *Y* are not d-separated given ***Z*** in *G*, then *X* and *Y* are dependent given ***Z*** in some distribution *P* that factorizes over *G*.

We use *I(G)* to denote the set of independencies that correspond to d-separation:
$I(G) = \{ (\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z}) : \text{d-sep}_G(\mathbf{X} ; \mathbf{Y} \mid \mathbf{Z})\}$.

I(G) = the set of independencies guaranteed in all P_G

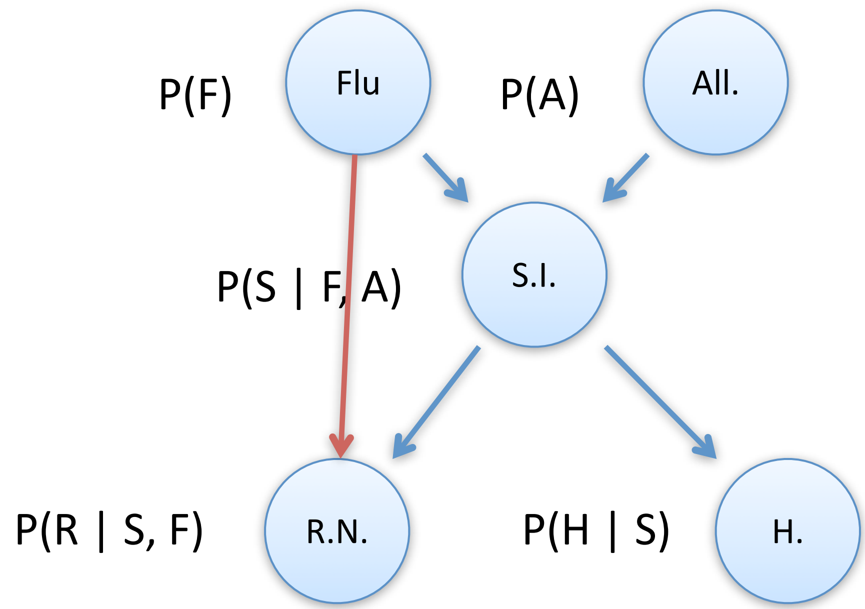
What's Independent?

- $F \perp A \mid \varnothing$
- $A \perp F \mid \varnothing$
- S?
- $R \perp \{F, A, H\} \mid S$
- $H \perp \{F, A, R\} \mid S$



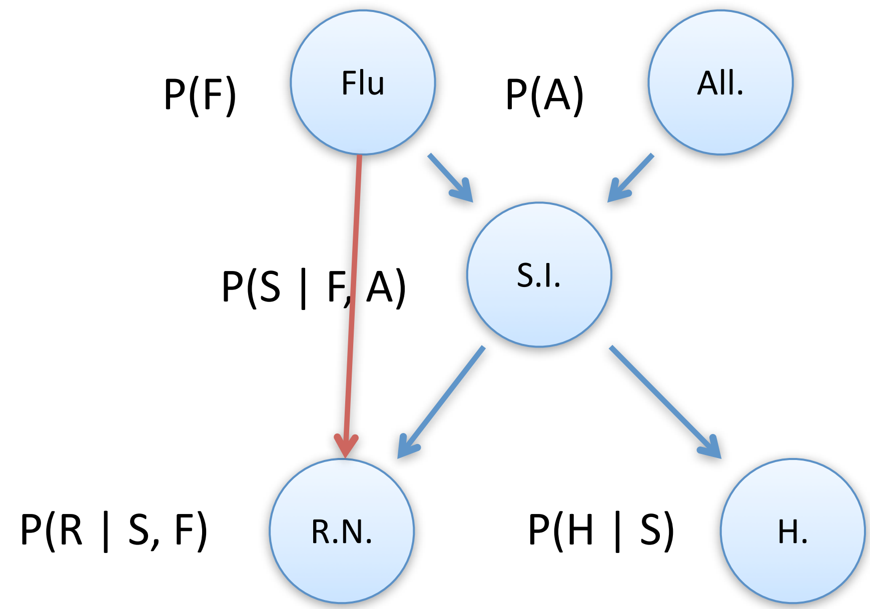
New Edge: What's Independent?

- $F \perp A \mid \varnothing$
- $A \perp F \mid \varnothing$
- S?
- $R \perp \{$~~F,~~ $A, H\} \mid S, F$
- $H \perp \{F, A, R\} \mid S$



New Edge: What's Independent?

- $F \perp A \mid \varnothing$
- $A \perp F \mid \varnothing$
- S?
- R ⊥ {~~F,~~ A, H} | S, F
- $H \perp \{F, A, R\} \mid S$
- $F \perp A \mid H$?



Questions

1. Given a BN,
   what distributions can be represented?
2. Given a distribution,
   what BNs can represent it?
3. In addition to the Local Markov Assumption,
   what other independence assumptions are encoded in a given BN?

Reality vs. Model

- World: true distribution P
  - true independencies
  - true factored form (beyond chain rule)

- Model: Bayesian network
  - a graph encoding local independence assumptions

- Any connections?

Representation Theorem

The conditional independencies in our BN are a subset of the independencies in P.

⇔

$$P(\boldsymbol{X}) = \prod_{i=1}^{n} P(X_i \mid \mathbf{Parents}(X_i))$$

- Given a graph G, can find I(G).
- Given a distribution P, can find I(P) (in theory anyway)
- **I-Map**: I(G) ⊂ I(P)
- **I-Equivalence**: $I(G_1) = I(G_2)$

I-Equivalence

- Two graphs G_1 and G_2 are **I-Equivalent** if $I(G_1) = I(G_2)$
- Define “Skeleton”: undirected version of G.

- Theorem 3.7. If G_1 and G_2 have the same skeleton and the same set of v-structures, then they are I-equivalent.

Minimal and Perfect I-Maps

- G is a **Minimal I-Map** for I if
  - G is an I-Map for I, and
  - the removal of any single edge would make it no longer an I-Map.
- G is a **P-Map** (Perfect Map) for I if
  - I(G) = I
- Is there a directed graphical model P-Map for every I?
  - No!

Homework #1

- Describe and discuss.