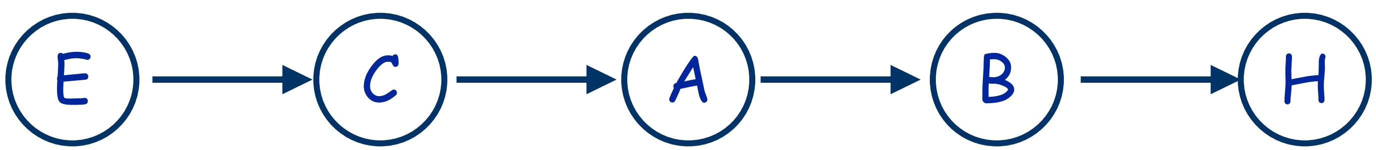
CSC384: Lecture 10

- Last time
  - Inference and Independence
- Today
  - Reasoning under uncertainty (belief networks)
- Readings:
  - Today: 10.3 (note: d-separation not covered in text)
  - Next week: 10.3 (var. elim.), 10.4 (decision making)

Exploiting Cond. Ind. (Recap)

- Let's see what conditional independence buys us
- Consider a story:
  - If Craig woke up too early E, Craig probably needs coffee C; if C, Craig needs coffee, he's likely angry A. If A, there is an increased chance of an aneurysm (burst blood vessel) B. If B, Craig is quite likely to be hospitalized H.



E – Craig woke too early A – Craig is angry H – Craig hospitalized
C – Craig needs coffee B – Craig burst a blood vessel

Cond’l Ind. in our Story (Recap)

- If you learned any of E, C, A, or B, your assessment of Pr(H) would change.
  - E.g., if any of these are seen to be true, you would increase Pr(h) and decrease Pr(~h).
  - So H is *not independent* of E, or C, or A, or B.
- But if you knew value of B (true or false), learning value of E, C, or A, would not influence Pr(H). Influence these factors have on H is mediated by their influence on B.
  - Craig doesn't get sent to the hospital because he's angry, he gets sent because he's had an aneurysm.
  - So H is *independent* of E, and C, and A, *given* B

Cond'l Ind. in our Story (Recap)

- So H is *independent* of E, and C, and A, *given* B
- Similarly:
  - B is *independent* of E, and C, *given* A
  - A is *independent* of E, *given* C
- This means that:
  - Pr(H | B, {A,C,E}) = Pr(H|B)
    - i.e., for any subset of {A,C,E}, this relation holds
  - Pr(B | A, {C,E}) = Pr(B | A)
  - Pr(A | C, {E}) = Pr(A | C)
  - Pr(C | E) and Pr(E) don't "simplify"

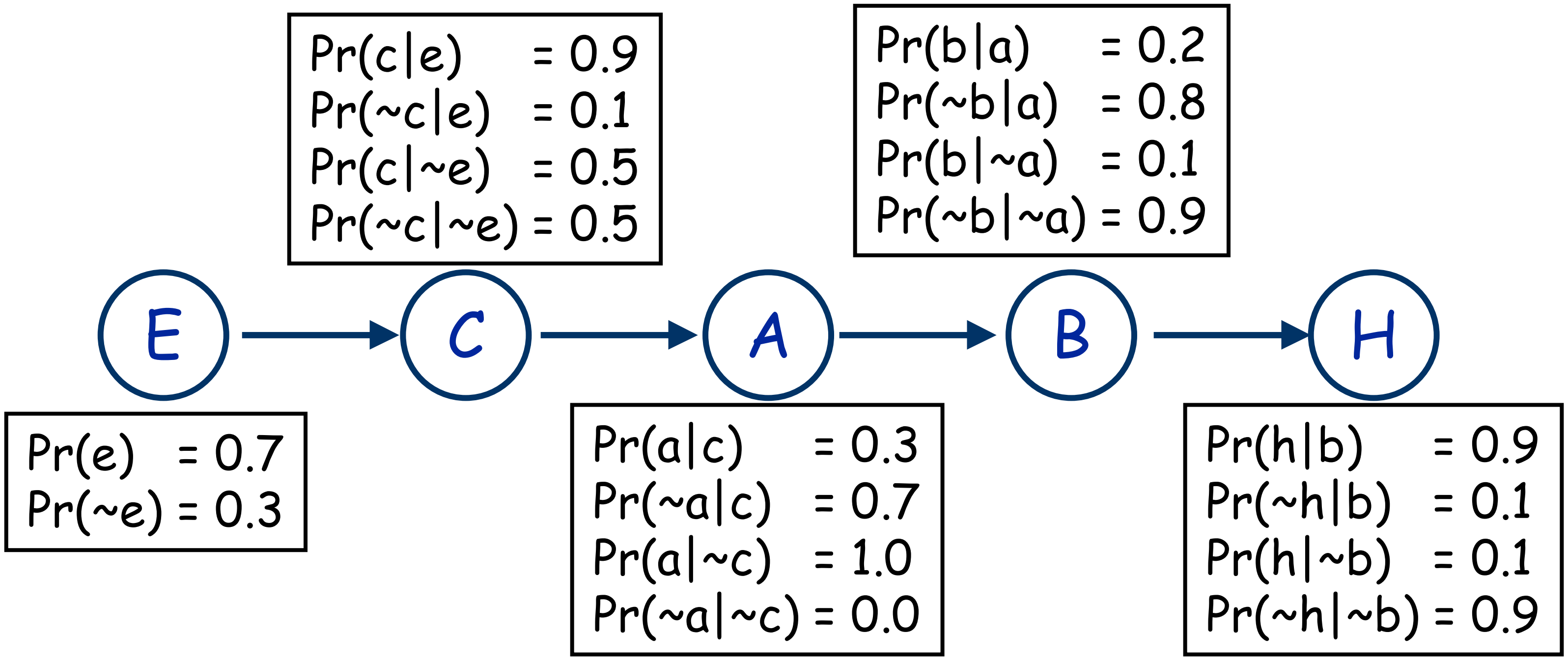
Cond'l Ind. in our Story (Recap)

- By the chain rule (for any instantiation of H…E):
  - Pr(H,B,A,C,E) =

    Pr(H|B,A,C,E) Pr(B|A,C,E) Pr(A|C,E) Pr(C|E) Pr(E)
- By our independence assumptions:
  - Pr(H,B,A,C,E) =

    Pr(H|B) Pr(B|A) Pr(A|C) Pr(C|E) Pr(E)
- We can specify the full joint by specifying five *local conditional distributions*: Pr(H|B); Pr(B|A); Pr(A|C); Pr(C|E); and Pr(E)

Example Quantification



- Specifying the joint requires only 9 parameters (if we note that half of these are “1 minus” the others), instead of 31 for explicit representation
  - linear in number of vars instead of exponential!
  - linear generally if dependence has a chain structure

Inference is Easy

- Want to know P(a)? Use summing out rule:

$$P(a) = \sum_{c_i \in Dom(C)} \Pr(a \mid c_i)\Pr(c_i)$$

$$= \sum_{c_i \in Dom(C)} \Pr(a \mid c_i) \sum_{e_i \in Dom(E)} \Pr(c_i \mid e_i)\Pr(e_i)$$

These are all terms specified in our local distributions!

Inference is Easy

- Computing P(a) in more concrete terms:
  - $P(c) = P(c|e)P(e) + P(c|{\sim}e)P({\sim}e)$

    $= 0.8 * 0.7 + 0.5 * 0.3 = 0.78$
  - $P({\sim}c) = P({\sim}c|e)P(e) + P({\sim}c|{\sim}e)P({\sim}e) = 0.22$
    - $P({\sim}c) = 1 - P(c)$, as well
  - $P(a) = P(a|c)P(c) + P(a|{\sim}c)P({\sim}c)$

    $= 0.7 * 0.78 + 0.0 * 0.22 = 0.546$
  - $P({\sim}a) = 1 - P(a) = 0.454$

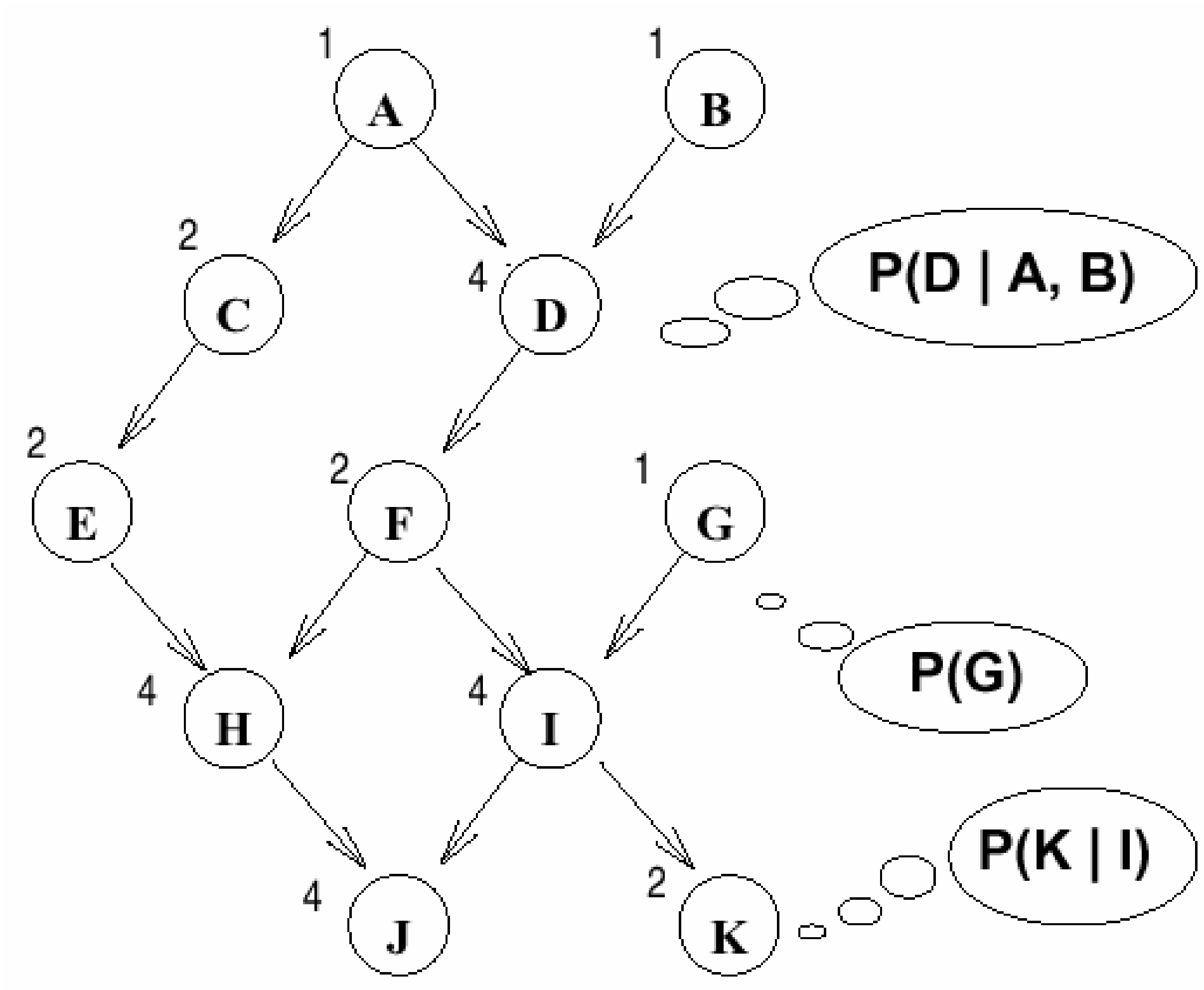
Bayesian Networks

- The structure above is a *Bayesian network*. A BN is a *graphical representation* of the direct dependencies over a set of variables, together with a set of *conditional probability tables (CPTs)* quantifying the strength of those influences.
- Bayes nets generalize the above ideas in very interesting ways, leading to effective means of representation and inference under uncertainty.

Bayesian Networks

- A BN over variables $\{X_1, X_2, \ldots, X_n\}$ consists of:
  - a DAG whose nodes are the variables
  - a set of CPTs $\Pr(X_i \mid Par(X_i))$ for each X_i
- Key notions (see text for defn's, all are intuitive):
  - parents of a node: $Par(X_i)$
  - children of node
  - descendents of a node
  - ancestors of a node
  - family: set of nodes consisting of X_i and its parents
    - CPTs are defined over families in the BN

An Example Bayes Net



- A couple CPTS are "shown"
- Explict joint requires 2^{11} -1 =2047 parmtrs
- BN requires only 27 parmtrs (the number of entries for each CPT is listed)

Semantics of a Bayes Net

- The structure of the BN means: every X_i is *conditionally independent of all of its nondescendants given it parents*:

$$\Pr(X_i \mid S \cup Par(X_i)) = \Pr(X_i \mid Par(X_i))$$

for any subset $S \subseteq NonDescendents(X_i)$

Semantics of Bayes Nets (2)

- If we ask for $Pr(x_1, x_2, \ldots, x_n)$ we obtain
  - assuming an ordering consistent with network

- By the chain rule, we have:

$$Pr(x_1, x_2, \ldots, x_n)$$
$$= Pr(x_n \mid x_{n-1}, \ldots, x_1)\, Pr(x_{n-1} \mid x_{n-2}, \ldots, x_1) \ldots Pr(x_1)$$
$$= Pr(x_n \mid Par(X_{n-1}))\, Pr(x_{n-1} \mid Par(x_{n-2})) \ldots Pr(x_1)$$

- Thus, the joint is recoverable using the parameters (CPTs) specified in an arbitrary BN

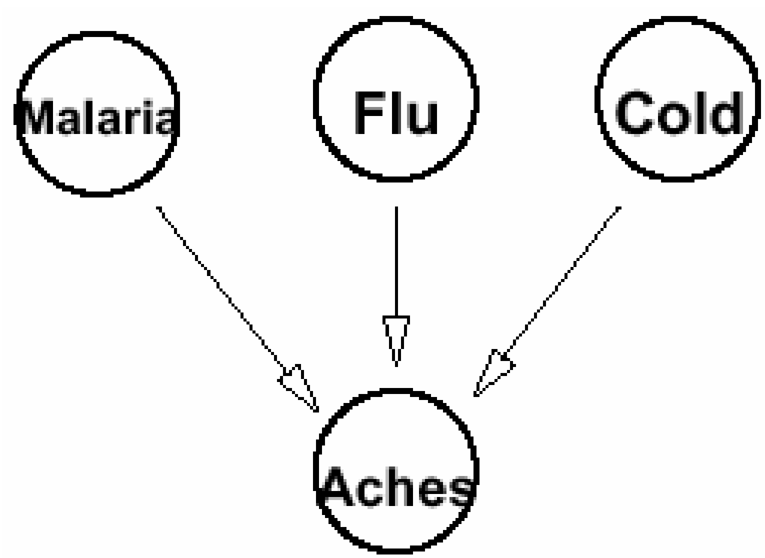
Constructing a Bayes Net

- Given any distribution over variables $X_1, X_2, \ldots, X_n$, we can construct a Bayes net that faithfully represents that distribution.

Take any ordering of the variables (say, the order given), and go through the following procedure for X_n down to X_1. Let Par(X_n) be any subset S $\subseteq$ $\{X_1, \ldots, X_{n-1}\}$ such that X_n is independent of $\{X_1, \ldots, X_{n-1}\}$ - S given S. Such a subset must exist (convince yourself). Then determine the parents of X_{n-1} the same way, finding a similar S $\subseteq$ $\{X_1, \ldots, X_{n-2}\}$, and so on. In the end, a DAG is produced and the BN semantics must hold by construction.

Causal Intuitions

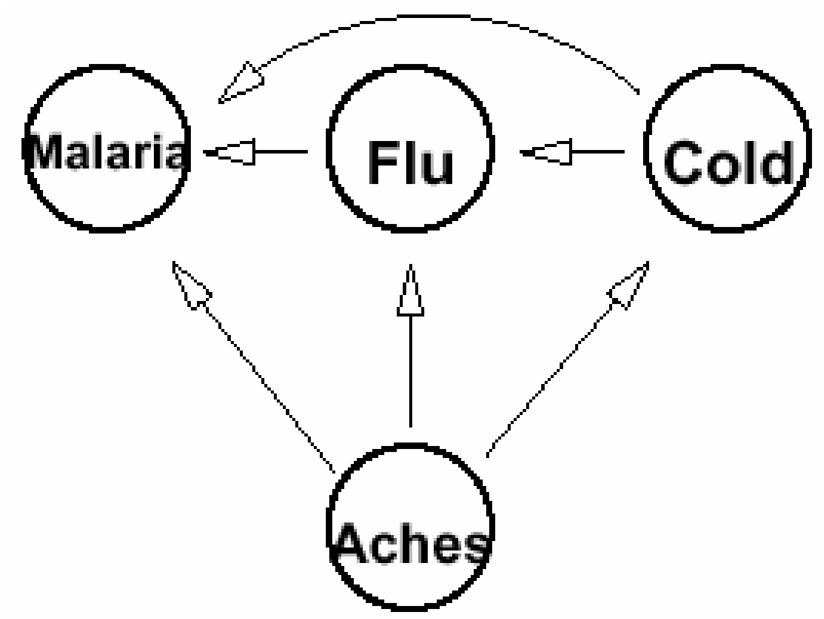
- The construction of a BN is simple
  - works with arbitrary orderings of variable set
  - but some orderings much better than others!
  - generally, if ordering/dependence structure reflects causal intuitions, a more natural, compact BN results



- In this BN, we've used the ordering Mal, Cold, Flu, Aches to build BN for distribution P
  - Variable can only have parents that come earlier in the ordering

Causal Intuitions

- Suppose we build the BN for distribution P using the opposite ordering
  - i.e., we use ordering Aches, Cold, Flu, Malaria
  - resulting network is more complicated!



- Mal depends on Aches; but it also depends on Cold, Flu *given* Aches
  - Cold, Flu explain away Mal given Aches
- Flu depends on Aches; but also on Cold *given* Aches
- Cold depends on Aches

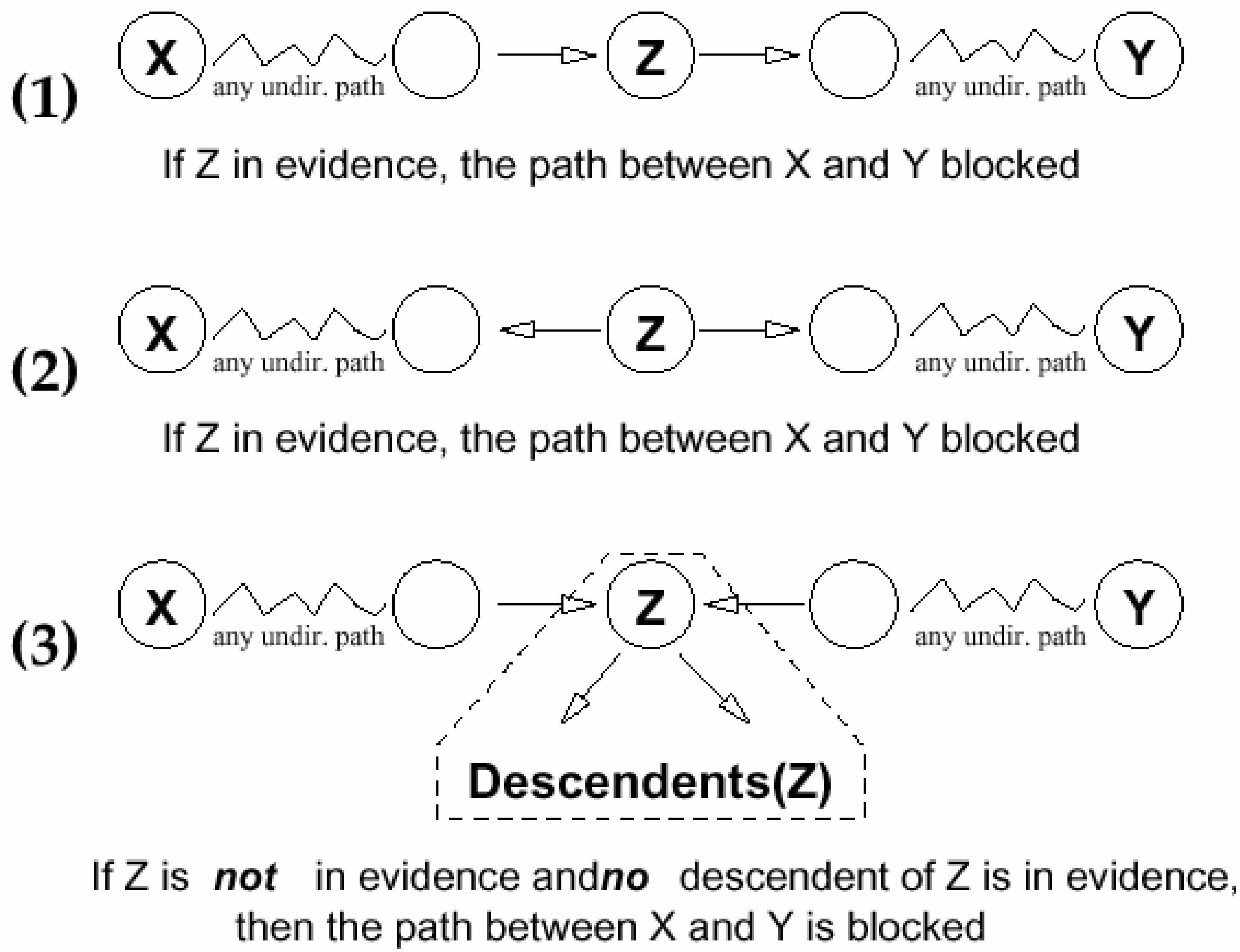
Testing Independence

- Given BN, how do we determine if two variables X, Y are independent (given evidence E)?
  - we use a (simple) graphical property
- **D-separation**: A set of variables **E** *d-separates* X and Y if it *blocks every undirected path* in the BN between X and Y. (We'll define *blocks* next.)
- X and Y are conditionally independent given evidence **E** if **E** d-separates X and Y
  - thus BN gives us an easy way to tell if two variables are independent (set E = ∅) or cond. independent

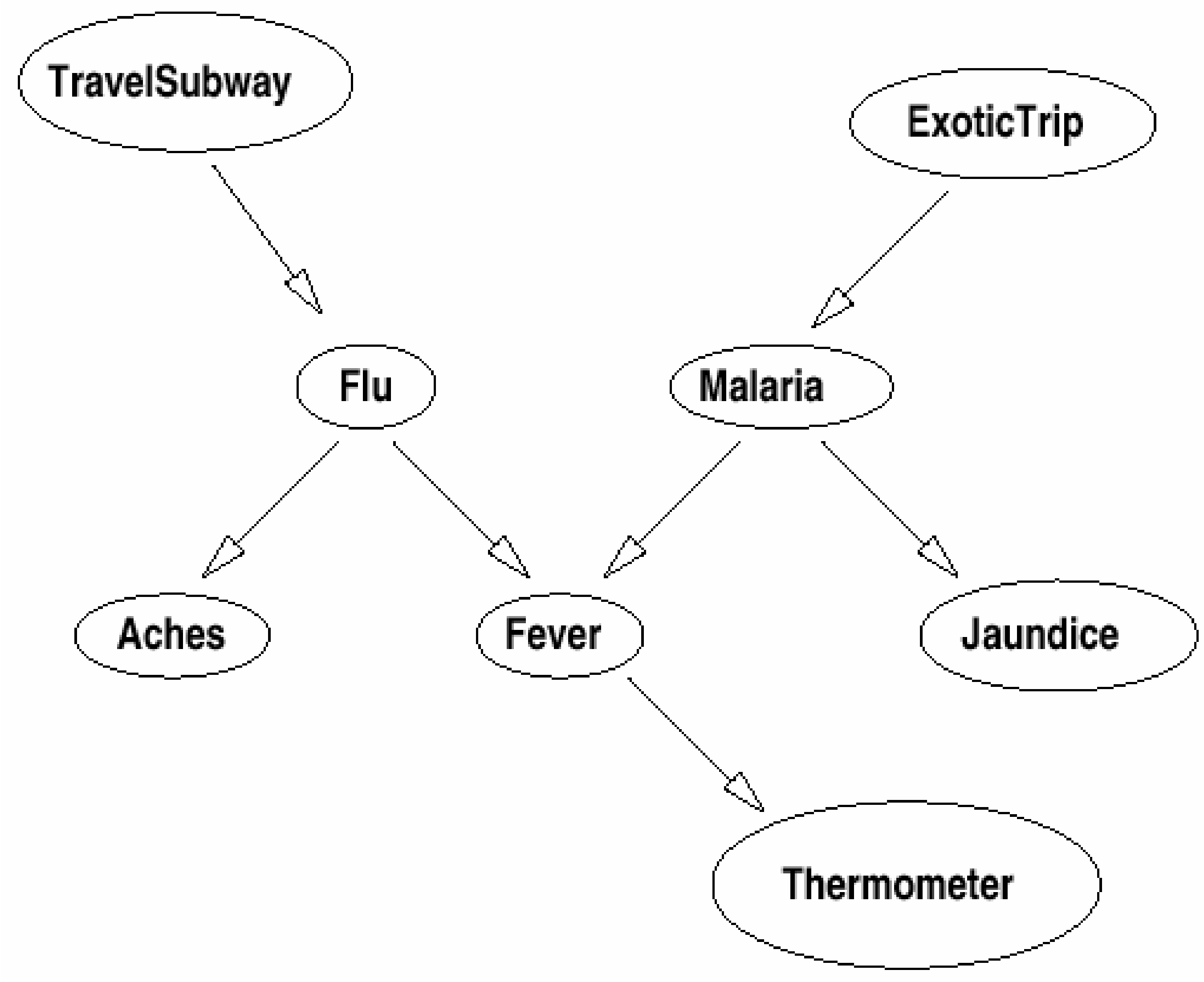
Blocking in D-Separation

- Let P be an undirected path from X to Y in a BN. Let **E** be an evidence set. We say **E** *blocks path P* iff there is some node Z on the path such that:
  - **Case 1:** one arc on P *goes into* Z and one *goes out* of Z, and Z∈**E**; or
  - **Case 2:** both arcs on P leave Z, and Z∈**E**; or
  - **Case 3:** both arcs on P enter Z and *neither Z, nor any of its descendents*, are in **E**.

Blocking: Graphical View



D-Separation: Intuitions



D-Separation: Intuitions

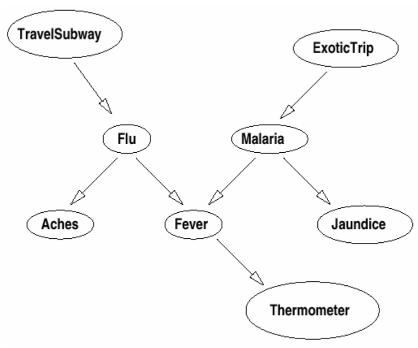
- Subway and Therm are dependent; but are independent given Flu (since Flu blocks the only path)
- Aches and Fever are dependent; but are independent given Flu (since Flu blocks the only path). Similarly for Aches and Therm (dependent, but indep. given Flu).
- Flu and Mal are indep. (given no evidence): Fever blocks the path, since it is *not in evidence*, nor is its decsendant Therm. Flu,Mal are dependent given Fever (or given Therm): nothing blocks path now.
- Subway,ExoticTrip are indep.; they are dependent given Therm; they are indep. given Therm and Malaria. This for exactly the same reasons for Flu/Mal above.

Inference in Bayes Nets

- The independence sanctioned by D-separation allows us to compute prior and posterior probabilities quite effectively.
- We'll look at a couple simple examples to illustrate. We'll focus on networks without *loops*. (A loop is a cycle in the underlying *undirected* graph. Recall the directed graph has no cycles.)

Simple Forward Inference (Chain)

- Computing prior require simple forward "propagation" of probabilities (using Subway net)



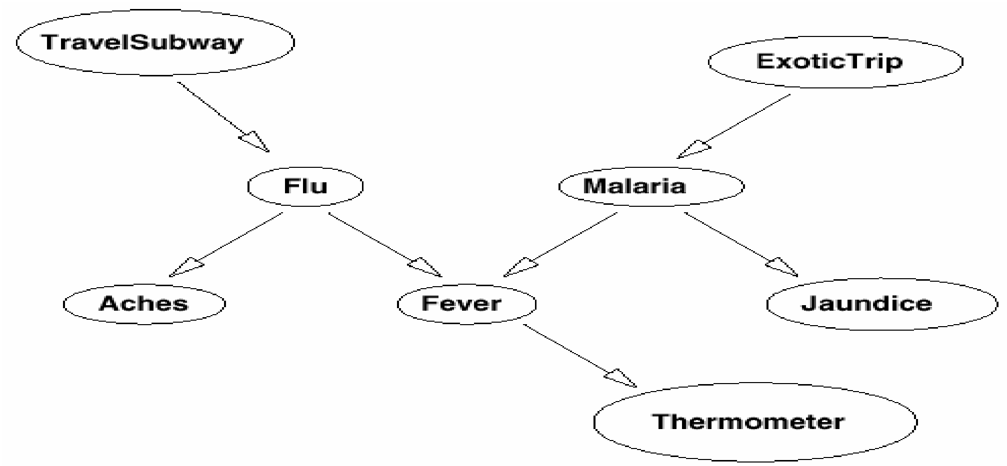
$$P(J) = \Sigma_{M,ET}\, P(J|M,ET)\, P(M,ET)$$
$$= \Sigma_{M,ET}\, P(J|M)\, P(M|ET)\, P(ET)$$
$$= \Sigma_{M}\, P(J|M)\, \Sigma_{ET}\, P(M|ET)\, P(ET)$$

- (1) follows by summing out rule; (2) by chain rule and independence; (3) by distribution of sum
  - Note: all (final) terms are CPTs in the BN
  - Note: only ancestors of J considered

Simple Forward Inference (Chain)

- Same idea applies when we have “upstream” evidence

$$P(J \mid et) = \Sigma_M P(J \mid M,et)\, P(M \mid et)$$
$$= \Sigma_M P(J \mid M)\, P(M \mid et)$$



Simple Forward Inference (Pooling)

- Same idea applies with multiple parents

$$P(Fev) = \Sigma_{Flu,M}\ P(Fev|Flu,M)\ P(Flu,M)$$
$$= \Sigma_{Flu,M}\ P(Fev|Flu,M)\ P(Flu)\ P(M)$$
$$= \Sigma_{Flu,M}\ P(Fev|Flu,M)\ \Sigma_{TS}\ P(Flu|TS)\ P(TS)\ \Sigma_{ET}\ P(M|ET)\ P(ET)$$

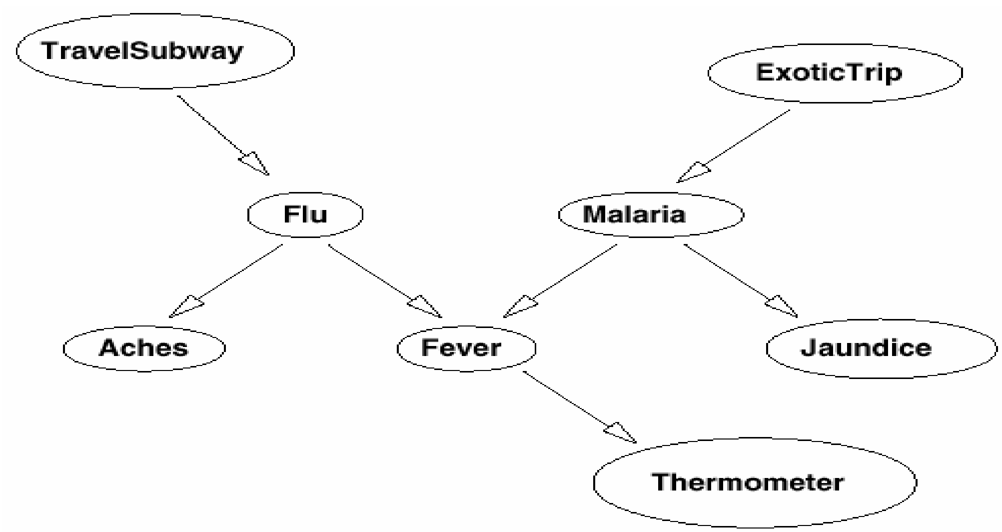
- (1) follows by summing out rule; (2) by independence of Flu, M; (3) by summing out
  - note: all terms are CPTs in the Bayes net

Simple Forward Inference (Pooling)

- Same idea applies with evidence

$$P(Fev|ts,\sim m) = \Sigma_{Flu}\, P(Fev\, |Flu,ts,\sim m)\, P(Flu|\, ts,\sim m)$$

$$= \Sigma_{Flu}\, P(Fev|Flu,\sim m)\, P(Flu|ts)$$



Simple Backward Inference

- When evidence is downstream of query variable, we must reason "backwards." This requires the use of Bayes rule:

$$P(ET \mid j) = \alpha P(j \mid ET) P(ET)$$
$$= \alpha \Sigma_M P(j \mid M,ET) P(M|ET) P(ET)$$
$$= \alpha \Sigma_M P(j \mid M) P(M|ET) P(ET)$$

- First step is just Bayes rule
  - normalizing constant α is 1/P(j); but we needn't compute it explicitly if we compute P(ET | j) for each value of ET: we just add up terms P(j | ET) P(ET) for all values of ET (they sum to P(j))

Backward Inference (Pooling)

- Same ideas when several pieces of evidence lie "downstream"

$$P(ET \mid j,fev) = P(j,fev \mid ET)\, P(ET)$$

$$= \Sigma_M P(j,fev \mid M,ET)\, P(M|ET)\, P(ET)$$

$$= \Sigma_M P(j,fev \mid M)\, P(M|ET)\, P(ET)$$

$$= \Sigma_M P(j \mid M)\, P(fev \mid M)\, P(M|ET)\, P(ET)$$

- Same steps as before; but now we compute prob of both pieces of evidence given hypothesis ET and combine them. Note: they are independent given M; but not given ET.
- Still must simplify P(fev|M) down to CPTs (as usual)

Variable Elimination

- The intuitions in the above examples give us a simple inference algorithm for networks without loops: the *polytree* algorithm. We won't discuss it further. But be comfortable with the intuitions.
- Instead we'll look at a more general algorithm that works for general BNs; but the propagation algorithm will more or less be a special case.
- The algorithm, ***variable elimination***, simply applies the summing out rule repeatedly. But to keep computation simple, it exploits the independence in the network and the ability to distribute sums inward.