Lecture 20

Bayesian Networks: Construction

Lecture Overview

- Recap lecture 19
- Bayesian networks: construction

- Defining Conditional Probabilities in a Bnet
- Considerations on Network Structure (time permitting)

Chain Rule

- Allows representing a Join Probability Distribution (JPD) as the product of conditional probability distributions

Theorem: Chain Rule

$$nn$$

$$PP(ff_1 \wedge \ldots \wedge ff_{nn}) = PP(ffff | ff_{ff-1} \wedge \ldots \wedge ff_1)$$

$ii=1$

E.g. $P(A,B,C,D) = P(A) \times P(B|A) \times P(C|A,B) \times P(D|A,B,C)$

Chain Rule example

$$P(f_1 \wedge \dots \wedge f_n) = \overset{n}{\underset{ii=1}{}} P(f f | f_{f\,-\,1\,\wedge\,\dots\,\wedge} f_1)$$

P(A,B,C,D)

= P(D|A,B,C) × P(A,B,C) =

= P(D|A,B,C) × P(C|A,B) × P(A,B)

= P(D|A,B,C) × P(C|B,A) × P(B|A) × P(A)

= P(A)P(B|A)P(C|A,B)P(D|A,B,C)

Why does the chain rule help us?

We will see how, under specific circumstances (variables independence), this rule helps gain compactness

- We can represent the JPD as a product of marginal distributions
- We can simplify some terms when the variables involved are marginally independent or conditionally independent

Marginal Independence

Definition (Marginal independence)
Random variable X is (marginally) independent of random variable Y, written $X \perp\!\!\!\perp Y$, if for all $x \in \mathrm{dom}(X)$, $y_j \in \mathrm{dom}(Y)$ and $y_k \in \mathrm{dom}(Y)$, the following equation holds:

$$\begin{aligned} & P(X = x | Y = y_j) \\ = & P(X = x | Y = y_k) \\ = & P(X = x) \end{aligned}$$

- Intuitively: if $X \perp\!\!\!\perp Y$, then
- learning that Y=y does not change your belief in X

- and this is true for all values y that Y could take

- For example, weather is marginally independent of the result of a coin toss

Exploiting marginal independence

- Recall the product rule $p(X=x \wedge Y=y) = p(X=x \mid Y=y) \times p(Y=y)$

- If X and Y are marginally independent, $p(X=x \mid Y=y) = p(X=x)$

- Thus we have $p(X=x \wedge Y=y) = p(X=x) \times p(Y=y)$
- In distribution form $p(X,Y) = p(X) \times p(Y)$

- If $X_1, \ldots, X_n$ are marginally independent, then we can represent their JPD as a product of marginal distributions

$$P(X_1, \ldots, Xn) = \prod_{i=1}^{n} P(Xi)$$

- If all of $X_1, \ldots, X_n$ are Boolean, how many entries does the JPD $P(X_1, \ldots, Xn)$ have?
  - One entry for each possible world: 2^n

- How many entries would all the marginal distributions have combined?
  - Each of the n tables only has two entries $P(X_1 = \text{true})$ and $P(X_1 = \text{false})$
  - So, in total: 2n. Exponentially fewer than the JPD!
  - Exponentially fewer than the JPD!

Exploiting marginal independence

Given the binary variables A,B,C,D,

To specify P(A,B,C,D) one needs the JDP below

A	B	C	D	P(A,B,C
T	T	T	T	
T	T	T	F	
T	T	F	T	
T	T	F	F	
T	F	T	T	
T	F	T	F	
T	F	F	T	
T	F	F	F	
F	T	T	T	

To specify P(A)×P(B) ×P(C)×P(D) one needs the JDPs below

A	P(A)
T	
F	

B	P(B)
T	
F	

C	P(C)
T	

F	T	T	F	
F	T	F	T	
F	T	F	F	
F	F	T	T	
F	F	T	F	
F	F	F	T	
F	F	F	F	

F	

D	P(D)
T	
F	

Conditional Independence

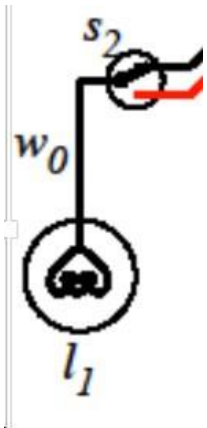
Definition (Conditional independence)
Random variable X is (conditionally) independent of random variable Y given random variable Z if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$, $y_k \in dom(Y)$ and $z_m \in dom(Z)$ the following equation holds:

$$P(X = xi|Y = y_j, Z = zm)$$
$$= P(X = xi|Y = y_k, Z = zm)$$
$$= P(X = xi|Z = zm)$$

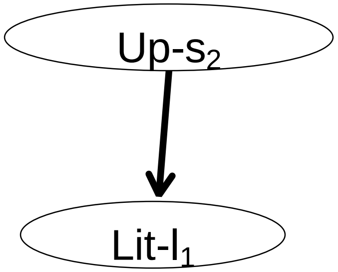
- Intuitively: if X and Y are conditionally independent given Z, then
- learning that Y=y does not change your belief in X when we already know Z=z

- and this is true for all values y that Y could take and all values z that Z could take

Example for Conditional Independence

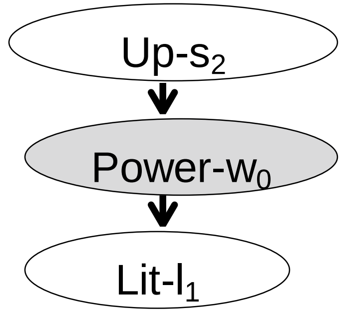


- Whether light l_1 is lit (Lit-l_1) and the position of switch s_2 (Up-s_2) are not marginally independent
- The position of the switch determines whether there is power in the wire w_0 connected to the light



- However, whether light l_1 is lit is conditionally independent from the position of switch s_2 given whether there is power in wire w_0 (Power-w_0)
- Once we know Power-w_0, learning values for Up-s_2 does not change our beliefs about Lit-l_1

- I.e., Lit-l_1 is conditionally independent of Up-s_2 given Power-w_0



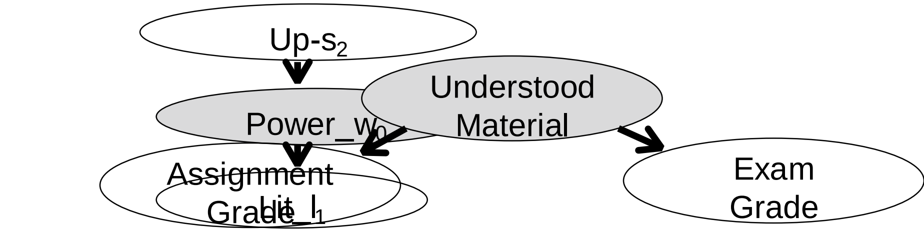
Conditional vs. Marginal Independence

Two variables can be

Conditionally but not marginally independent

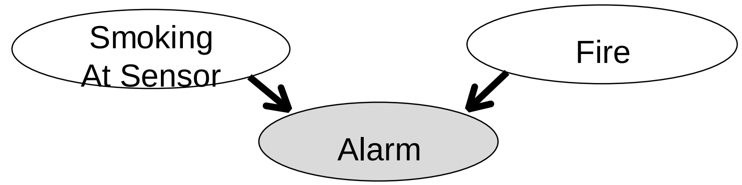
- ExamGrade and AssignmentGrade
- ExamGrade and AssignmentGrade given UnderstoodMaterial

- Lit-l1 and Up-s2
- Lit-l1 and Up-s2 given Power_w_0



Marginally but not conditionally independent

- SmokingAtSensor and Fire
- SmokingAtSensor and Fire given Alarm

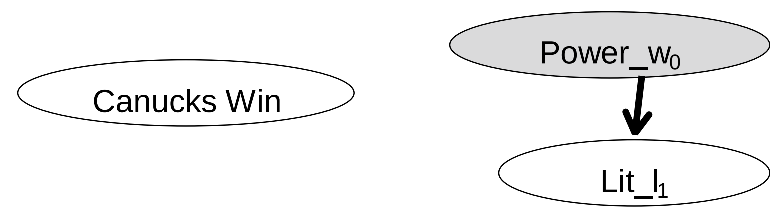


Both marginally and conditionally independent

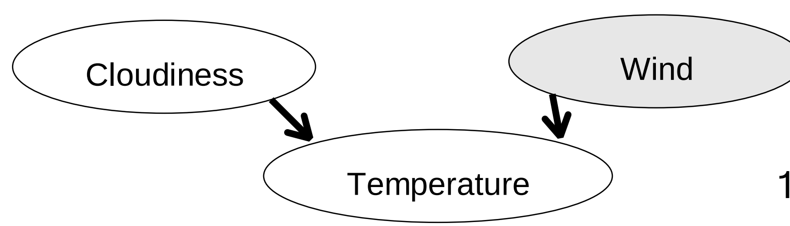
- CanucksWinStanleyCup and Lit_l_1
- CanucksWinStanleyCup and Lit_l_1 given Power_w

Neither marginally nor conditionally independent

- Temperature and Cloudiness
- Temperature and Cloudiness given Wind



0



Exploiting Conditional Independence

Example 2: Boolean variables A,B,C,D

- D is conditionally independent of both A and B given C
  - ✓We can rewrite P(D | A,B,C) as P(D|C)

- P(D|C) is much simpler to specify than P(D | A,B,C) !

Definition (Conditional independence)

Random variable X is (conditionally) independent of random variable Y given random variable Z if, for all $x_i \in dom(X)$, $y_j \in dom(Y)$, $y_k \in dom(Y)$ and $z_m \in dom(Z)$ the following equation holds:

$$\begin{aligned} & P(X = xi \mid Y = y_j, Z = zm) \\ = \ & P(X = xi \mid Y = y_k, Z = zm) \\ = \ & P(X = xi \mid Z = zm) \end{aligned}$$

If A, B, C, D are Boolean variables

P(D | A,B,C) is given by the following table

A	B	C	P(D=T|A,B,C)	P(D=F|A,B,C)
T	T	T		
T	T	F		
T	F	T		
T	F	F		
F	T	T		
F	T	F		
F	F	T		
F	F	F		

8 - each row represents the probability distribution for D given the values that A, B and C take in that row

P(D|C) is given by the following table

C	P(D=T|C)	P(D=F|C)
T		

2 - each row represents the probability distribution for D given the value that C takes in that row

Putting It All Together

- Given the JPD P(A,B,C,D), we can apply the chain rule to get

$$P(A, B, C, D) = P(A) \times P(B \mid A) \times P(C \mid A, B) \times P(D \mid A, B, C)$$

- If D is conditionally independent of A and B given C, we can rewrite the above as

$$P(A, B, C, D) = P(A) \times P(B \mid A) \times P(C \mid A, B) \times P(D \mid C)$$

Under independence we gain compactness (fewer/smaller distributions to deal with)

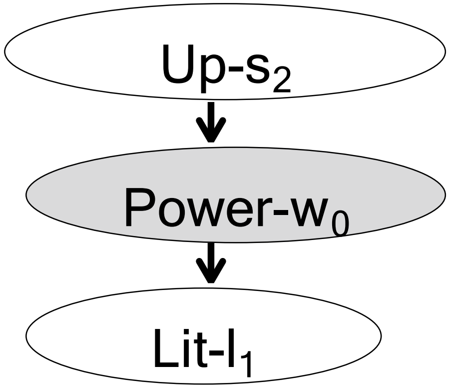
- The chain rule allows us to write the JPD as a product of conditional distributions
- Conditional independence allows us to write them more compactly

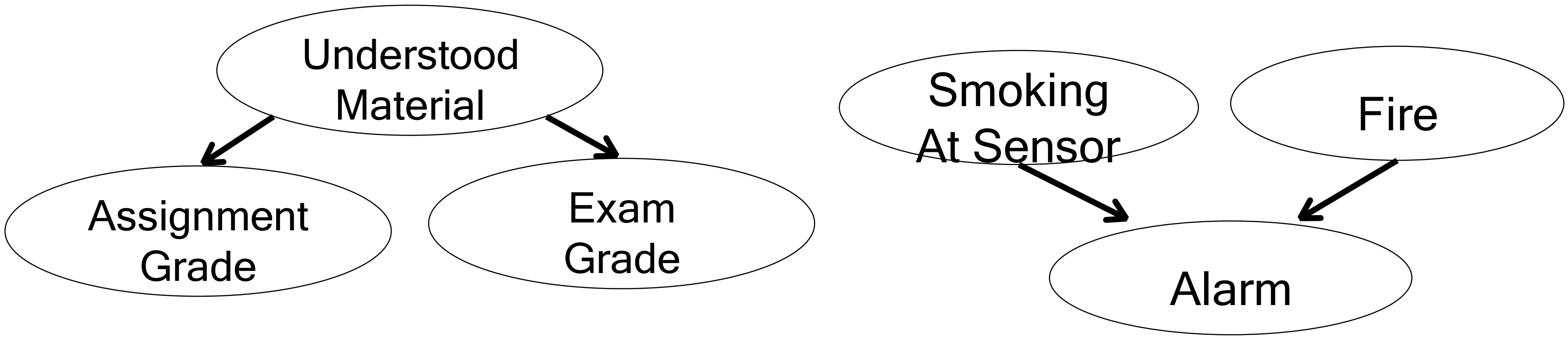
Bayesian (or Belief) Networks

- Bayesian networks and their extensions are Representation & Reasoning systems explicitly defined to exploit independence in probabilistic reasoning

Bayesian Networks: Intuition

- A graphical representation for a joint probability distribution
- Nodes are random variables
- Directed edges between nodes reflect dependence
- Some informal examples:





Belief (or Bayesian) networks

Def. A Belief network consists of

- a directed, acyclic graph (DAG) where each node is associated with a random variable X_i
- A domain for each variable X_i
- a set of conditional probability distributions for each node X_i given its parents $Pa(X_i)$ in the graph

$$P\ (X_i\ |\ Pa(X_i))$$

- The parents $Pa(X_i)$ of a variable X_i are those X_i directly depends on
- A Bayesian network is a compact representation of the JDP for a set of variables $(X_1, \ldots, X_n)$

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i))$$

Lecture Overview

- Recap lecture 19
- Bayesian networks: construction
- Defining Conditional Probabilities in a Bnet

- Considerations on Network Structure (time permitting)

How to build a Bayesian network

1. Define a total order over the random variables: $(X_1, \ldots, X_n)$

2. Apply the chain rule

   $$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1})$$

   Predecessors of X_i in the total order defined over the variables

3. For each X_i, select the smallest set of predecessors $Pa(X_i)$ such that

   X_i is conditionally independent from all its

$P(X_i \mid X_1, \ldots, X_{i-1}) = P(X_i \mid Pa(X_i))$ other predecessors given $Pa(X_i)$

4. Then we can rewrite

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i))$$

- This is a compact representation of the initial JPD • factorization of the JPD based on existing conditional independencies among the variables

How to build a Bayesian network (cont'd)

5. Construct the Bayesian Net (BN)

- Nodes are the random variables
- Draw a directed arc from each variable in $Pa(X_i)$ to X_i

- Define a conditional probability table (CPT) for each variable X_i:
- $P(X_i \mid Pa(X_i))$

Example for BN construction: Fire Diagnosis

You want to diagnose whether there is a fire in a building

- You can receive reports (possibly noisy) about whether everyone is leaving the building
- If everyone is leaving, this may have been caused by a fire alarm
- If there is a fire alarm, it may have been caused by a fire or by tampering
- If there is a fire, there may be smoke

Start by choosing the random variables for this domain, here all are Boolean:

- Tampering (T) is true when the alarm has been tampered with

- Fire (F) is true when there is a fire
- Alarm (A) is true when there is an alarm
- Smoke (S) is true when there is smoke
- Leaving (L) is true if there are lots of people leaving the building
- Report (R) is true if the sensor reports that lots of people are leaving the building

Next apply the procedure described earlier

Example for BN construction: Fire Diagnosis

1. Define a total ordering of variables:
   - Let's choose an order that follows the causal sequence of events
   - Fire (F), Tampering (T), Alarm, (A), Smoke (S) Leaving (L) Report (R)
2. Apply the chain rule

$P(F,T,A,S,L,R) =$

Example for BN construction: Fire Diagnosis

1. Define a total ordering of variables:
   - Let's chose an order that follows the causal sequence of events
   - Fire (F), Tampering (T), Alarm, (A), Smoke (S) Leaving (L) Report (R)
2. Apply the chain rule

P(F,T,A,S,L,R) =

P(F)P (T | F) P (A | F,T) P (S | F,T,A) P (L | F,T,A,S) P (R | F,T,A,S,L)

We will do steps 3, 4 and 5 together, for each element $P(X_i \mid X_1, \ldots, X_{i-1})$ of the factorization

3. For each variable (X_i), choose the parents Parents(X_i)by evaluating conditional independencies, so that

$$P(X_i \mid X_1, \dots, X_{i-1}) = P(X_i \mid \text{Parents}(X_i))$$

4. Rewrite

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i \mid \text{Parents}(X_i))$$

5. Construct the Bayesian network

Fire Diagnosis Example

P(F)P (T | F) P (A | F,T) P (S | F,T,A) P (L | F,T,A,S) P (R | F,T,A,S,L)

Fire (F) is the first variable in the ordering, X_1. It does not have parents.

Example

P(F)P (T | F) P (A | F,T) P (S | F,T,A) P (L | F,T,A,S) P (R | F,T,A,S,L)

- Tampering (T) is independent of fire (learning that one is true/false would not change your beliefs about the probability of the other)

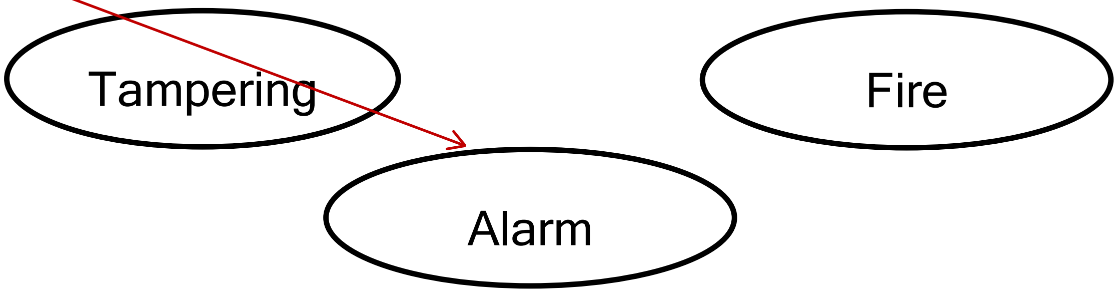
Example

$P(F)P(T)\,P(A \mid F,T)\,P(S \mid F,T,A)\,P(L \mid F,T,A,S)\,P(R \mid F,T,A,S,L)$

- Tampering (T) is independent of fire (learning that one is true/false would not change your beliefs about the probability of the other)

Fire Diagnosis Example

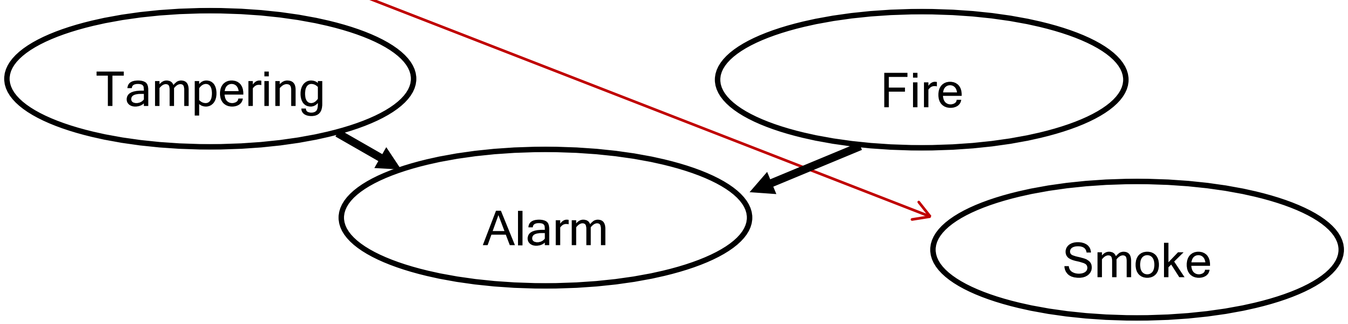
$P(F)P(T)\,P(A \mid F,T)\,P(S \mid F,T,A)\,P(L \mid F,T,A,S)\,P(R \mid F,T,A,S,L)$



- Alarm (A) depends on both Fire and Tampering: it could be caused by either or both

Fire Diagnosis Example

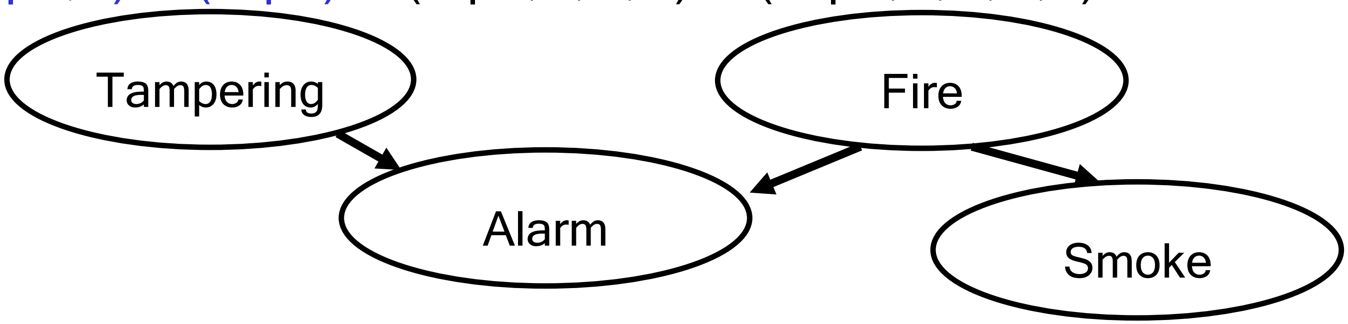
$P(F)P(T \mid F)\,P(A \mid F,T)\,P(S \mid F,T,A)\,P(L \mid F,T,A,S)\,P(R \mid F,T,A,S,L)$



Fire Diagnosis Example

- Smoke (S) is caused by Fire, and so is independent of Tampering and Alarm given whether there is a Fire

P(F)P (T | F) P (A | F,T) P (S | F) P (L | F,T,A,S) P (R | F,T,A,S,L)

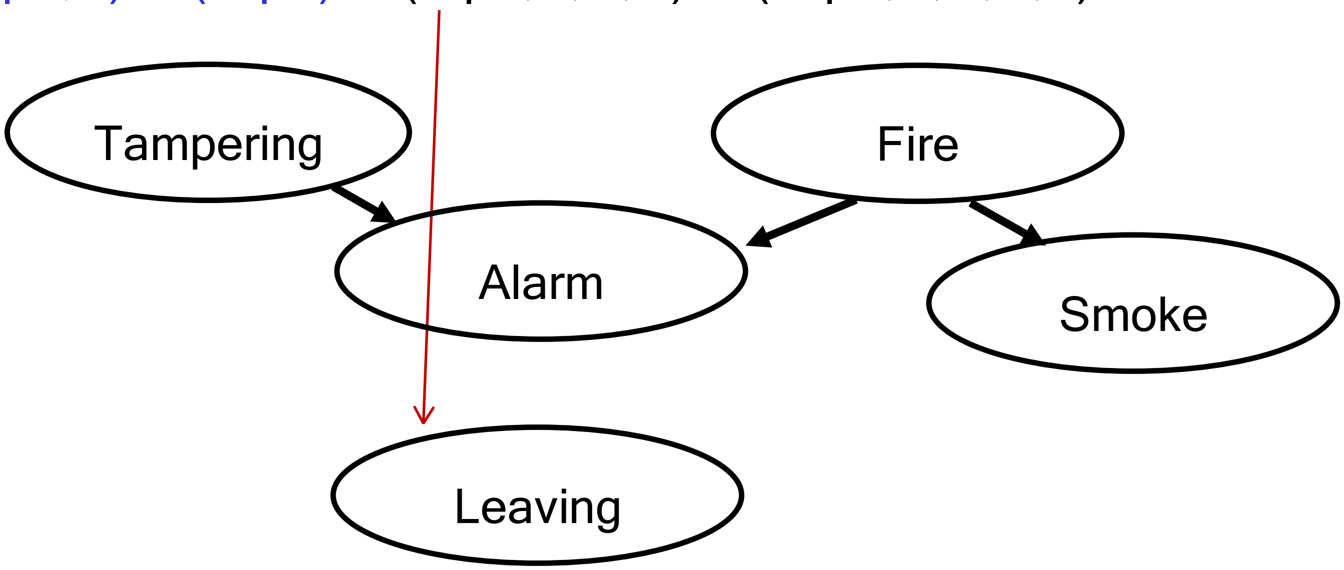


Fire Diagnosis Example

- Smoke (S) is caused by Fire, and so is independent of Tampering and Alarm given whether there is a Fire

Example

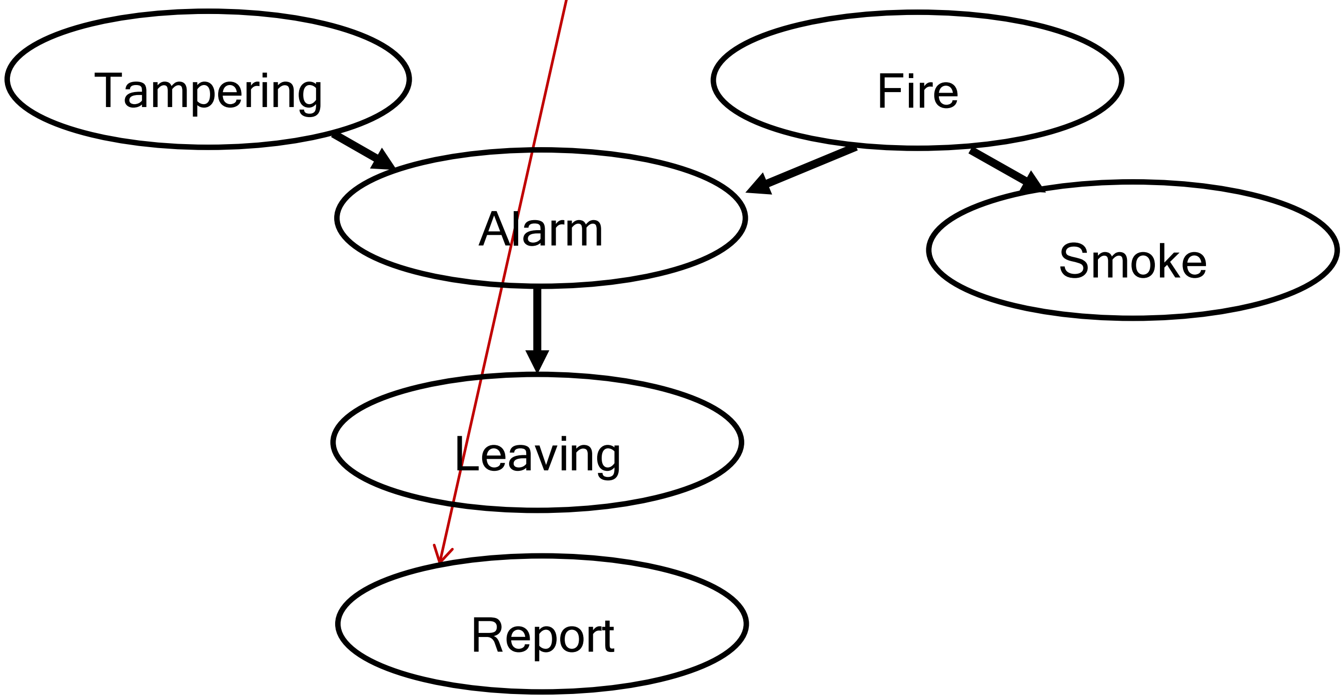
$P(F)P(T \mid F)\, P(A \mid F,T)\, P(S \mid F)\, P(L \mid F,T,A,S)\, P(R \mid F,T,A,S,L)$



- Leaving (L) is caused by Alarm, and thus is independent of the other variables given Alarm

Fire Diagnosis Example

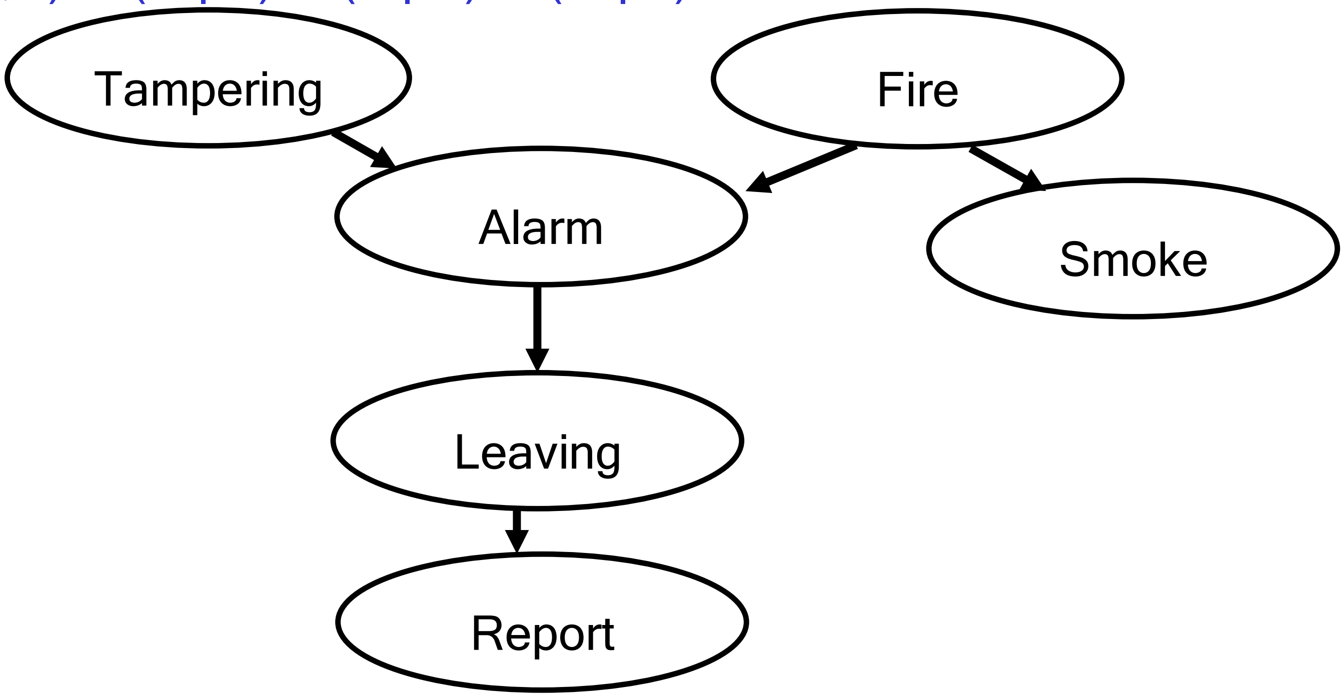
P(F)P (T) P (A | F,T) P (S | F) P (L | A) P (R | F,T,A,S,L)



- Report (R) is caused by Leaving, and thus is independent of the other variables given Leaving

Fire Diagnosis Example

P(F)P (T) P (A | F,T) P (S | F) P (L | A) P (R | L)



The result is the Bayesian network above, and its corresponding, very compact factorization of the original JPD

P(F,T,A,S,L,R)= P(F)P (T) P (A | F,T) P (S | F) P (L | A) P (R | L)

Example for BN construction: Fire Diagnosis

- Note that we intermixed steps 3, 4 and 5, just because sometime it is easier to reason about conditional dependencies graphically
- However, you can do step 3 and 4 first
- That this, you can simplify the product before building the network
- Still have to reason about dependencies between each node and its predecessors in the total order

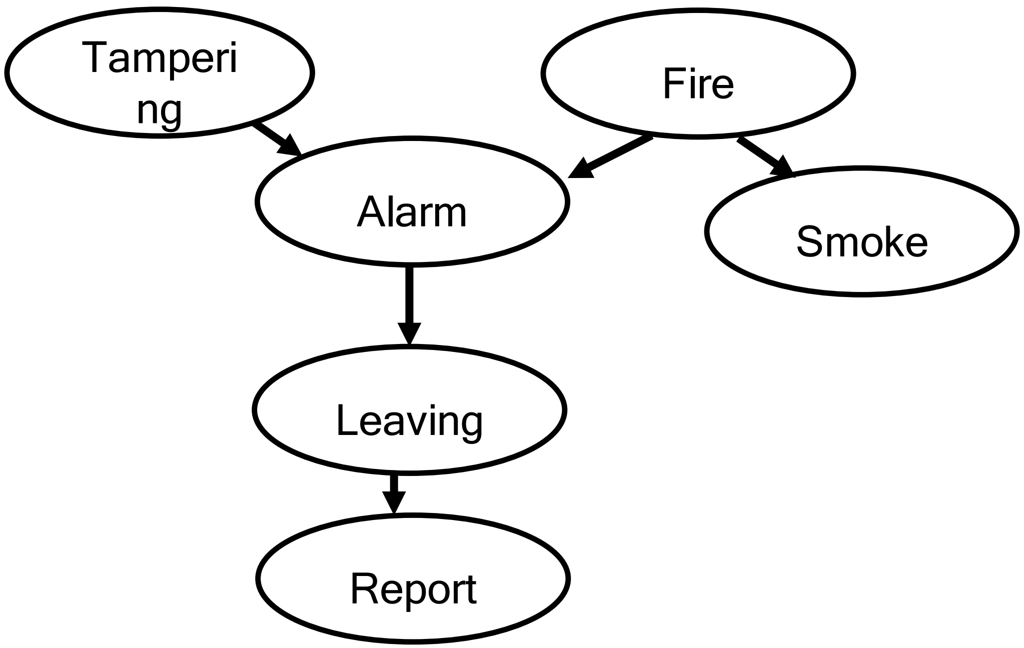
P(F)P (T | F) P (A | F,T) P (S | F,T,A) P (L | F,T,A,S) P (R | F,T,A,S,L)

Fire Diagnosis Example

$P(F)P(T)P(A \mid F,T)P(S \mid F)P(L \mid A)P(R \mid L)$

5. Construct the Bayesian Net (BN)

- Nodes are the random variables
- Draw a directed arc from each variable in Pa(X) to X
- Define a conditional probability table (CPT) for each variable X :
  - $P(X_i \mid Pa(X_i))$



Lecture Overview

- Recap lecture 19
- Bayesian networks: construction
- Defining Conditional Probabilities in a Bnet
- Considerations on Network Structure (time permitting)

Example for BN construction: Fire Diagnosis

- We are not done yet: must specify the Conditional Probability Table (CPT) for each variable. All variables are Boolean.
- How many probabilities do we need to specify for this Bayesian network?
- For instance, how many probabilities do we need to explicitly specify for Fire?

Example for BN construction: Fire Diagnosis

P(Fire=t)

A. 1 B. 2 C. 4 D. 8

- We are not done yet: must specify the Conditional Probability Table (CPT) for each variable. All variables are Boolean.
- How many probabilities do we need to specify for this Bayesian network?

Example for BN construction: Fire Diagnosis

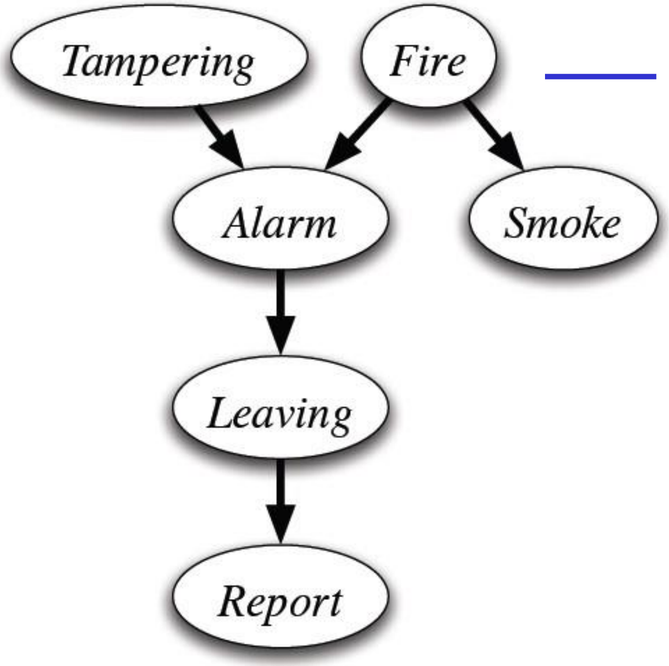
- For instance, how many probabilities

P(Fire): 1 probability -> P(Fire =T) 0.01
Because P(Fire = F) = 1 - P(Fire = T)

do we need to explicitly specify for Fire?

- How many probabilities do we need to explicitly specify for Alarm?

Example for BN construction: Fire Diagnosis

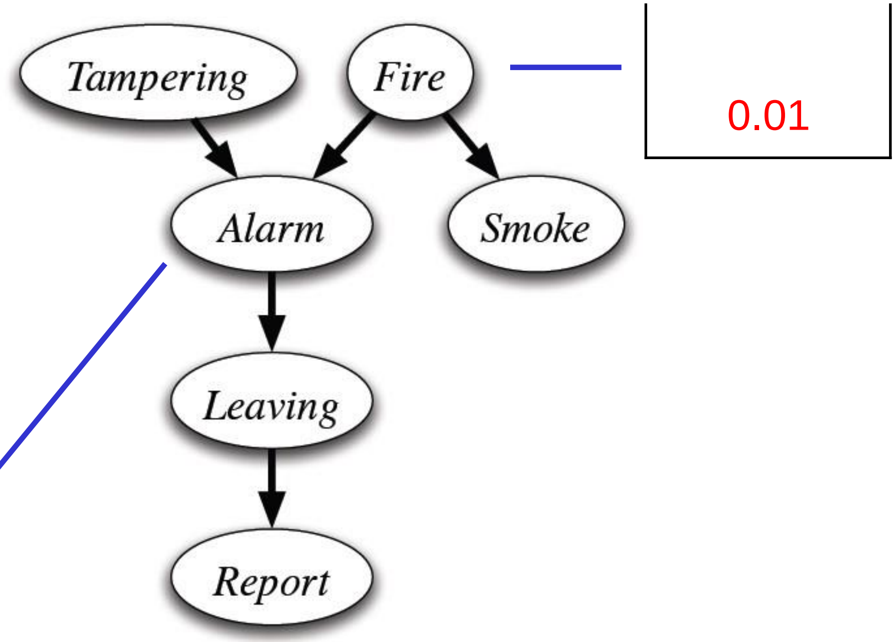


P(Fire=t)
0.01

- How many probabilities do we need to explicitly specify for Alarm?

P(Alarm|Tampering, Fire): 4 probabilities, 1 probability for each of the 4 instantiations of the parents

Example for BN construction: Fire Diagnosis

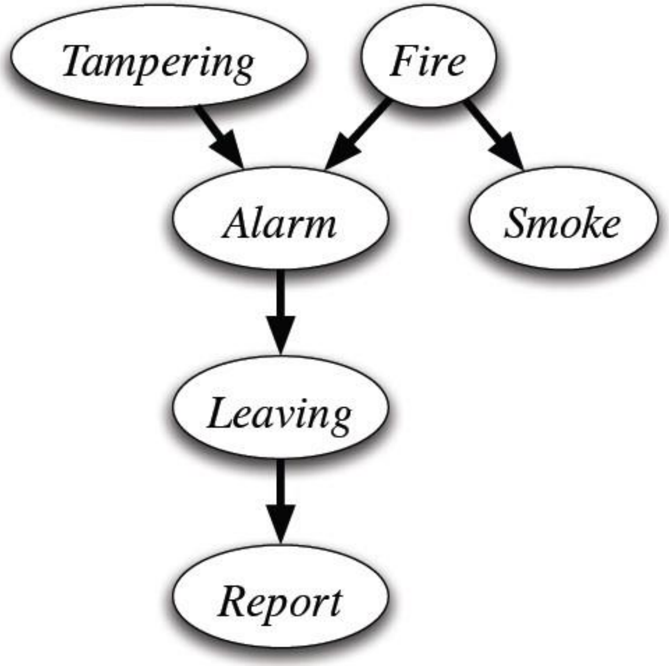


Tampering T	Fire F	P(Alarm=t|T,F)	P(Alarm=f|T,F)
t	t	0.5	0.5
t	f	0.85	0.15
f	t	0.99	0.01
f	f	0.0001	0.9999

We don't need to speficy explicitly P(Alarm=f|T,F) since probabilities in each row must sum to 1

Each row of this table is a conditional probability distribution

Example for BN construction: Fire Diagnosis



P(Fire=t)

- How many probabilities do we need to explicitly specify for the whole Bayesian network?

A. 6 B. 12 C. 20 D. 2^6-1

Example for BN construction: Fire Diagnosis

P(Tampering=t)
0.02

Tampering T	*Fire F*	*P(Alarm=t|T,F)*
t	t	0.5
t	f	0.85
f	t	0.99
f	f	0.0001

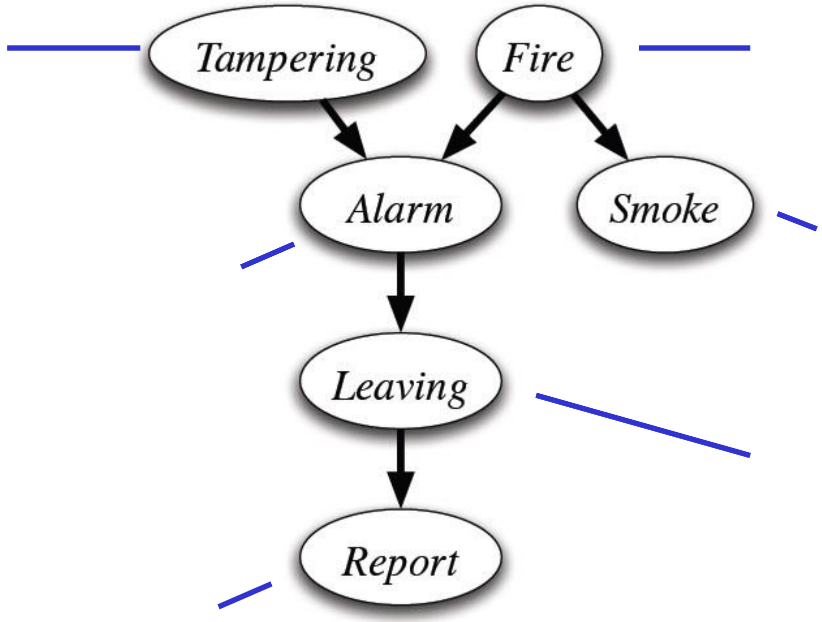
Leaving	*P(Report=t|L)*
t	0.75
f	0.01

P(Fire=t)
0.01

Fire F	*P(Smoke=t |F)*
t	0.9
f	0.01

Alarm	*P(Leaving=t|A)*
t	0.88
f	0.001

Example for BN construction: Fire Diagnosis



……..probabilities in total, compared to the of the JPD for P(T,F,A,S,L,R)

Leaving	*P(Report=t|L)*
t	0.75
f	0.01

Example for BN construction: Fire Diagnosis

12 probabilities in total, compared to the 2^6 - 1= 63 of the JPD for P(T,F,A,S,L,R)

P(Tampering=t)
0.02

P(Fire=t)
0.01

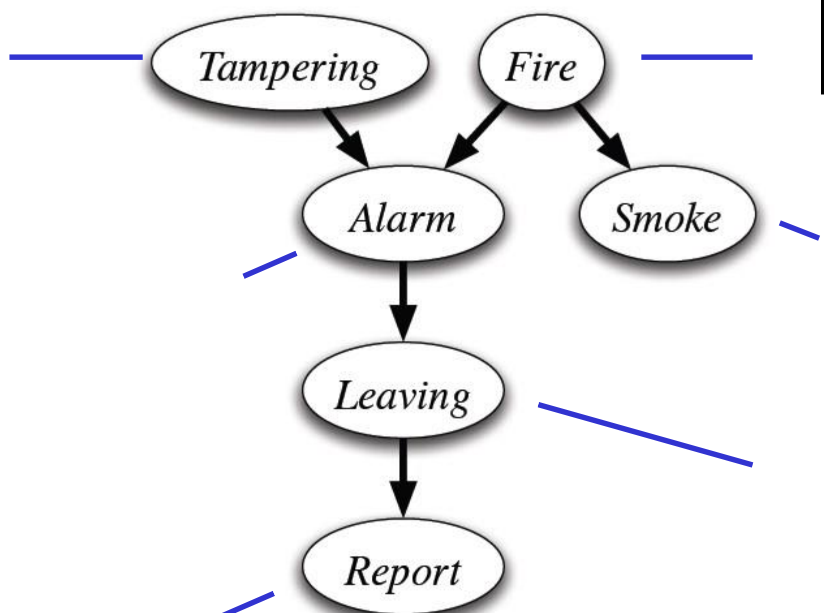
Tampering T	*Fire F*	*P(Alarm=t|T,F)*
t	t	0.5
t	f	0.85
f	t	0.99
f	f	0.0001

How many probabilities do we need to specify for this Bayesian network?

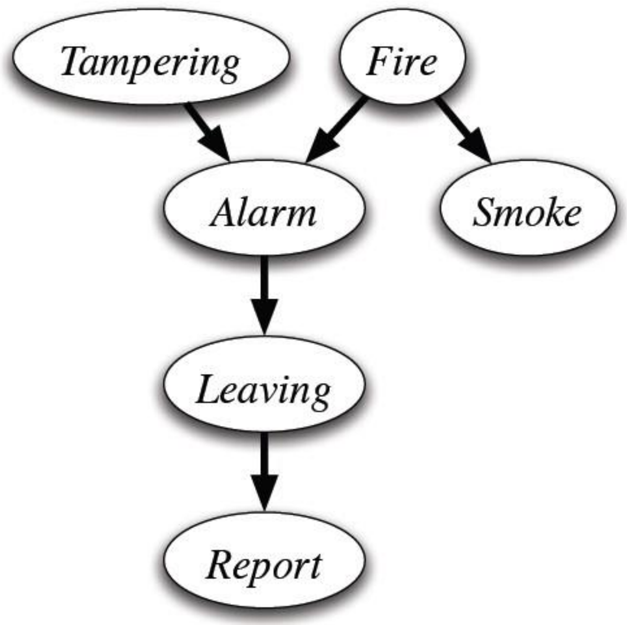
✓ P(Tampering): 1 probability P(T = t)

Fire F	*P(Smoke=t |F)*
t	0.9
f	0.01

Alarm	*P(Leaving=t|A)*
t	0.88
f	0.001



Example for BN construction: Fire Diagnosis



✓ P(Alarm|Tampering, Fire): 4 (independent)

- 1 probability for each of the 4 instantiations of the parents

✓ For all other variables with only one parent

- 2 probabilities: one for the parent being true and one for the parent being false

✓ In total: 1+1+4+2+2+2 = 12 (compared to 2^6-1= 63 for full JPD!)

Example for BN construction: Fire Diagnosis

Tampering T	*Fire F*	*P(Alarm=t|T,F)*
t	t	0.5
t	f	0.85
f	t	0.99
f	f	0.0001

Leaving	*P(Repo*
t	0.7
f	0.0

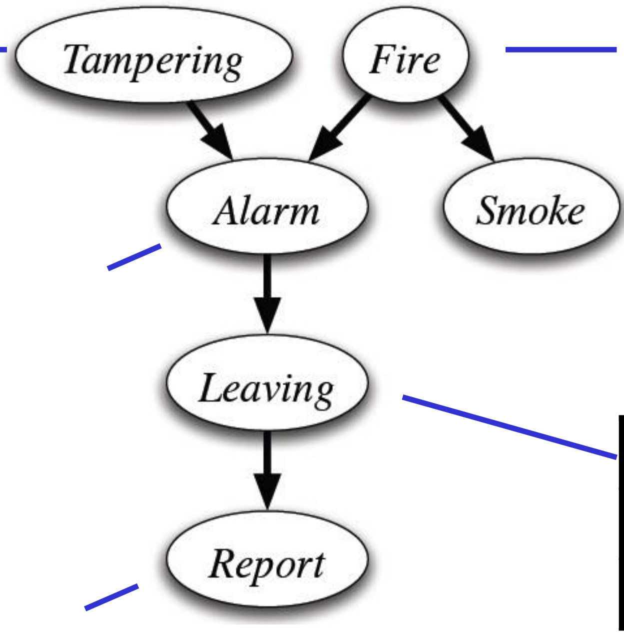
Once we have the CPTs in the network, we can compute any entry of the JPD

P(Tampering=t, Fire=f, Alarm=t, Smoke=f, Leaving=t, Report=t) =

Example for BN construction: Fire Diagnosis

P(Tampering=t)
0.02

P(Fire=t)
0.01



Fire F	P(Smoke=t |F)
t	0.9
f	0.01

Alarm	P(Leaving=t|A)
t	0.88
f	0.001

P(Tampering=t) x P(Fire=f)xP(Alarm=t| Tampering=t, Fire=f)xP(Smoke=f| Fire = f)xP(Leaving=t| Alarm=t) x P(Report=t|Leaving=t) =

= 0.02 x (1-0.01) x 0.85 x (1-0.01) x 0.88 x 0.75 = 0.126

In Summary

- In a Belief network, the JPD of the variables involved is defined as the product of the local conditional distributions

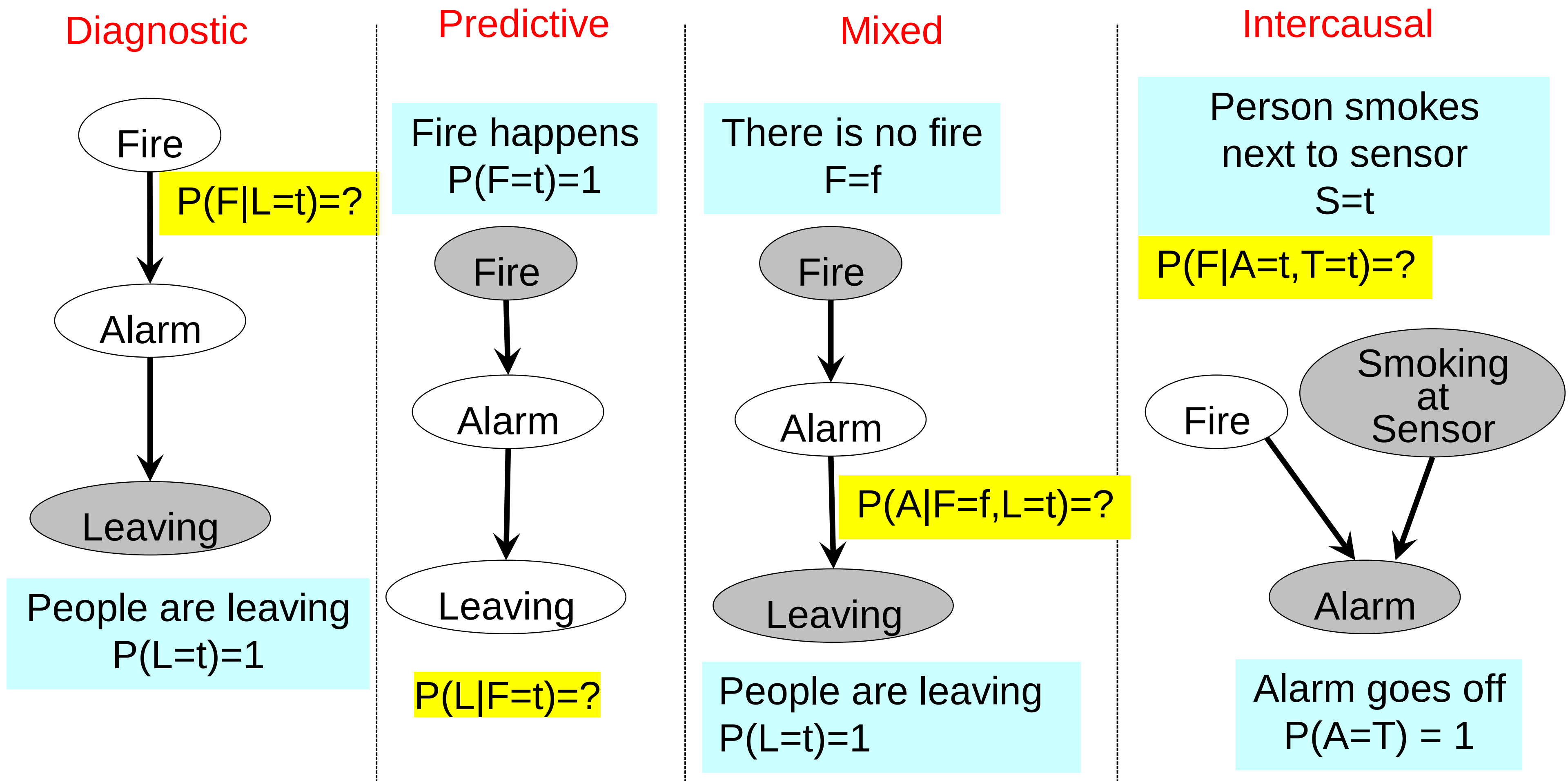
$$P(X_1, \dots, X_n) = \prod_i P(X_i \mid X_1, \dots, X_{i-1}) = \prod_i P(X_i \mid Parents(X_i))$$

- Any entry of the JPD can be computed given the

  CPTs in the network

Once we know the JPD, we can answer any query about any subset of the variables - (see Inference by Enumeration topic)

Thus, a Belief network allows one to answer any query on any subset of the variables



There are algorithms that leverage the Bnet structure to perform query answer efficiently

- For instance variable elimination, which we will cover soon
- First, however, we will think a bit more about network structure

Learning Goals so Far

- Given a JPD
- Marginalize over specific variables
- Compute distributions over any subset of the variables
- Use inference by enumeration
- to compute joint posterior probability distributions over any subset of variables given evidence
- Define and use marginal and conditional independence
- Build a Bayesian Network for a given domain (structure)
- Specify the necessary conditional probabilities

- Compute the representational savings in terms of number of probabilities required

Compactness

- In a Bnet, how many rows do we need to explicitly store for the CPT of a Boolean variable X_i with k Boolean parents?

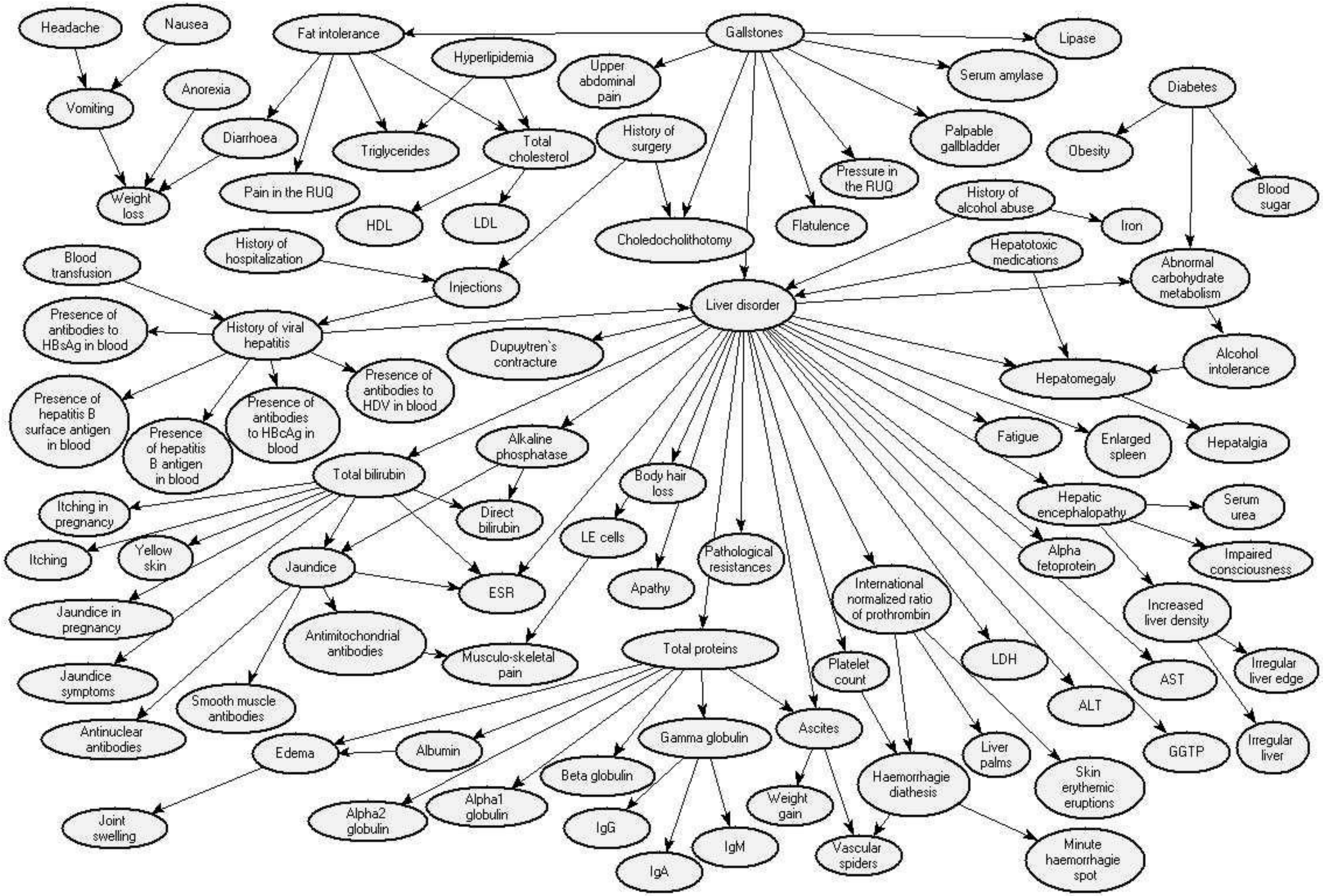
Compactness

- A CPT for a Boolean variable X_i with k Boolean parents has 2^k rows for the combinations of parent values
- If each variable has no more than k parents, the complete network requires to specify $n2^k$ numbers
- For $k \ll n$, this is a substantial improvement,
- the numbers required grow linearly with n, vs. $O(2^n)$ for the full joint distribution
- E.g., if we have a Bnets with 30 boolean variables, each with 5 parents

- Need to specify $30*2^5$ probability
- But we need 2^{30} for JPD

Realistic BNet: Liver Diagnosis

Source: Onisko et al., 1999



~ 60 nodes, max 4 parents per node

Need ~ $60 \times 2^4 = 15 \times 2^6$ probabilities instead of 2^{60} probabilities for the JPD

Compactness

- What happens if the network is fully connected?
- Or $k \approx n$
- Not much saving compared to the numbers needed to specify the full JPD
- Bnets are useful in sparse(or locally structured) domains
- Domains in with each component interacts with (is related to) a small fraction of other components

- What if this is not the case in a domain we need to reason about?

May need to make simplifying assumptions to reduce the dependencies in a domain

“Where do the numbers (CPTs) come from?”

From experts

- Tedious
- Costly
- Not always reliable

From data => Machine Learning

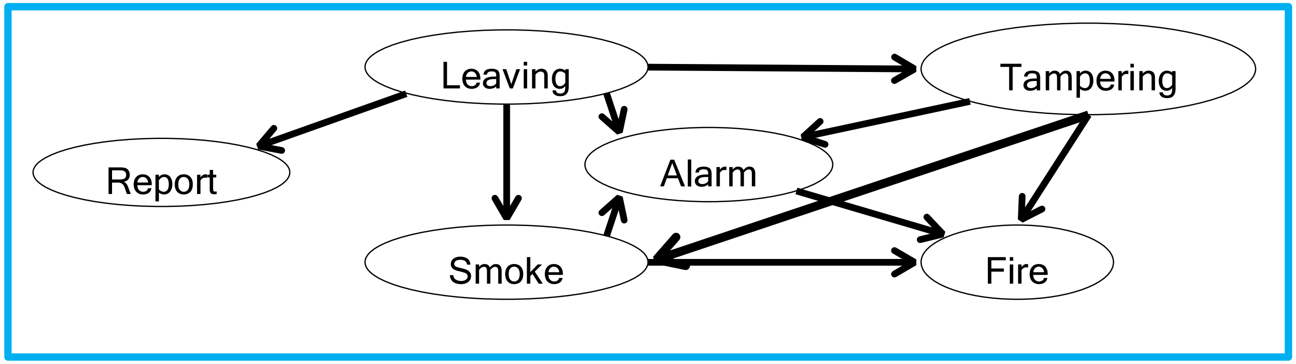
- There are algorithms to learn both structures and numbers (CPSC 340, CPSC 422)
- Can be hard to get enough data

Still, usually better than specifying the full JPD

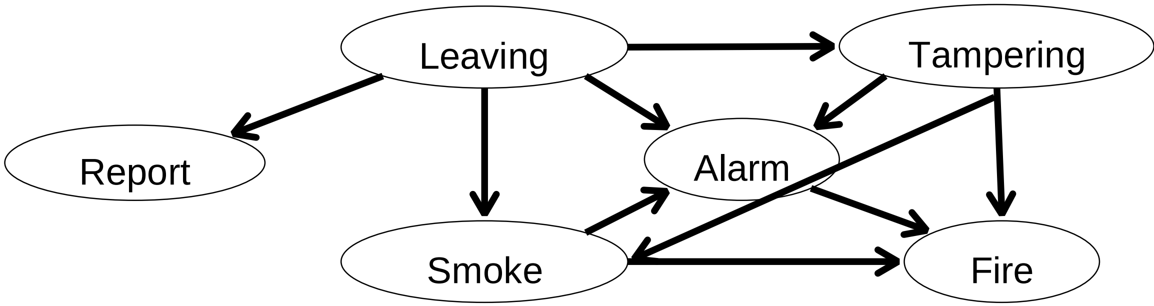
What if we use a different ordering?

- What happens if we use the following order:
- Leaving; Tampering; Report; Smoke; Alarm; Fire.

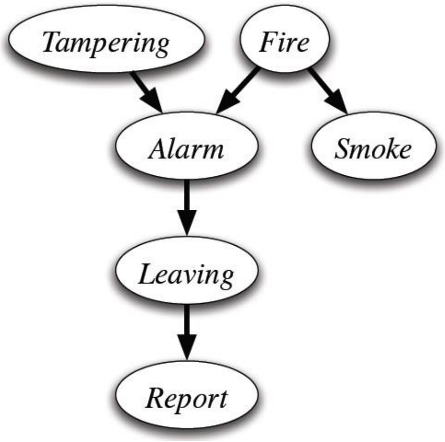
- We end up with a completely different network structure! (try it as an exercise)



- Which of the two structures is better?



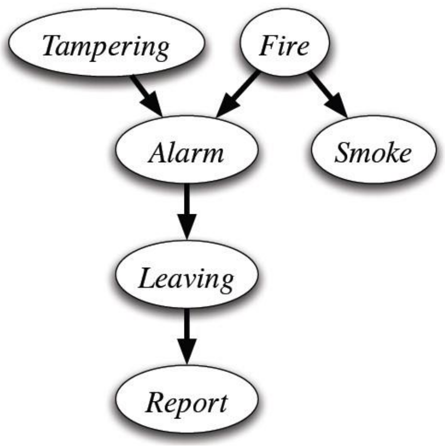
Which Structure is Better?

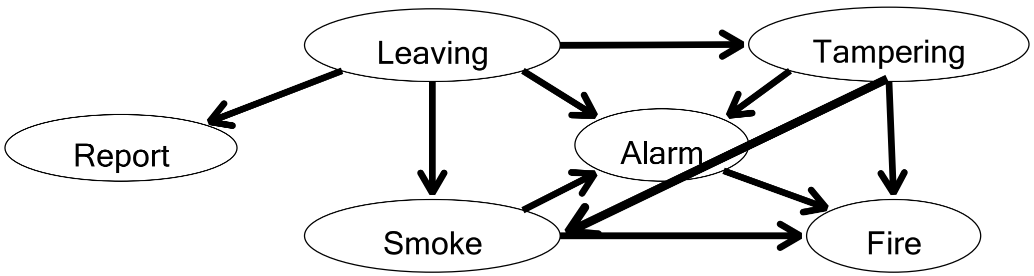


- Non-causal network is less compact: 1+2+2+4+8+8 = 25 numbers needed
- Deciding on conditional independence is hard in non-causal directions
    - Causal models and conditional independence seem hardwired for humans!
- Specifying the conditional probabilities may be harder than in causal direction
- For instance, we have lost the direct dependency between alarm and one of its causes, which essentially describes the alarm's reliability (info often provided by the maker)

Example contd.

- Other than that, our two Bnets for the Alarm problem are equivalent as long as they represent the same probability distribution





Variable ordering: L,T,R,S,A,F

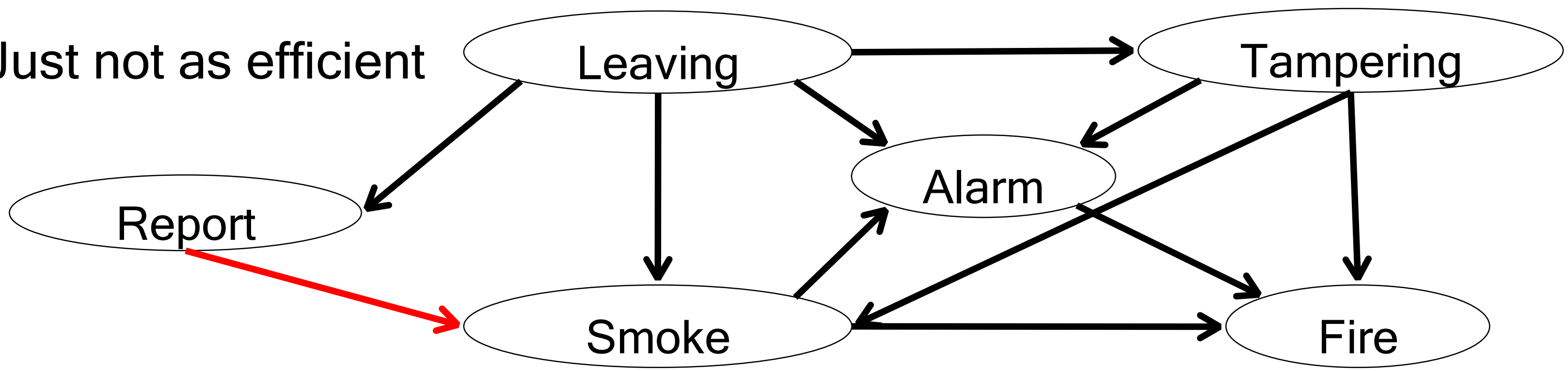
Variable ordering: T,F,A,S,L,R

P(T,F,A,S,L,R) = P (T) P (F) P (A | T,F) P (L | A) P (R|L) =

= P(L)P(T|L)P(R|L)P(S|L,T)P(A|S,L,T) P(F|S,A,T) i.e., they are equivalent if the corresponding CPTs are specified so that they satisfy the equation above

Are there wrong network structures?

- Given an order of variables, a network with arcs in excess to those required by the direct dependencies implied by that order are still ok
  - Just not as efficient



P (L)P(T|L)P(R|L) P(S|L,R,T) P(A|S,L,T) P(F|S,A,T) =
P (L)P(T|L)P(R|L)P(S|L,T)P(A|S,L,T) P(F|S,A,T)

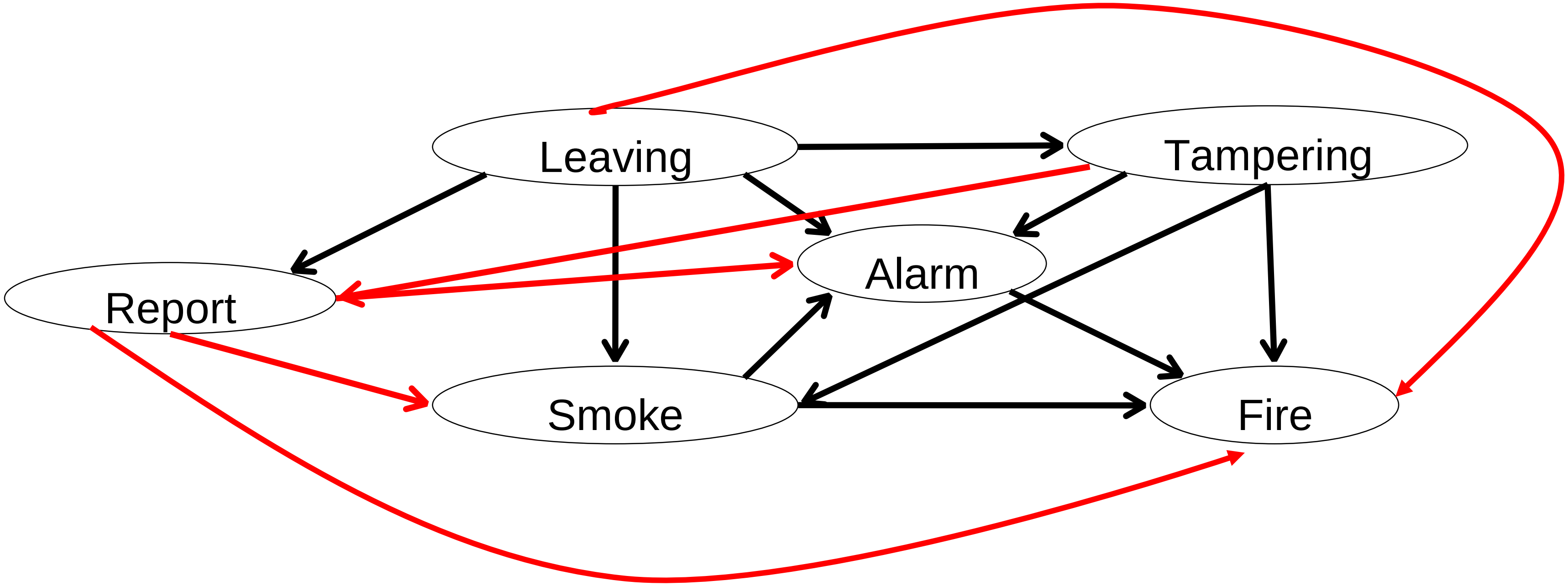
- One extreme: the fully connected network is always correct but rarely the best choice

Are there wrong network structures?

- It corresponds to just applying the chain rule to the JDP, without leveraging conditional independencies to simplify the factorization

P(L,T,R,S,A,L)= P (L)P(T|L)P(R|L,T)P(S|L,T,R)P(A|S,L,T,R) P(F|S,A,T,L,R)

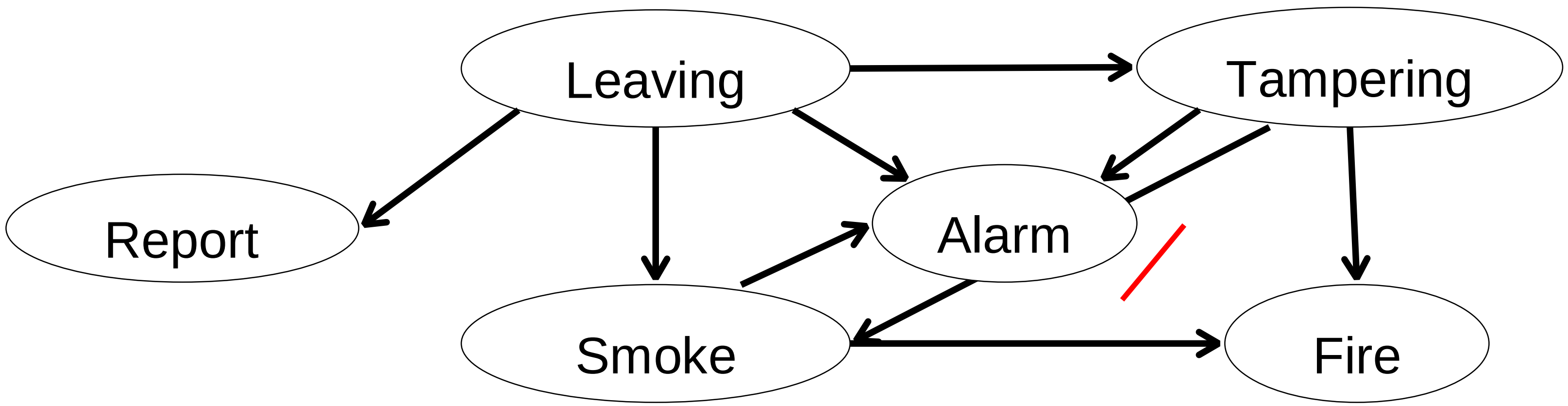
P(L,T,R,S,A,L)= P (L)P(T|L)P(R|L,T)P(S|L,T,R)P(A|S,L,T,R) P(F|S,A,T,L,R)



- How can a network structure be wrong?

Are there wrong network structures?

- If it misses directed edges that are required
- E.g. an edge is missing below, making Fire conditionally independent of Alarm given Tampering and Smoke

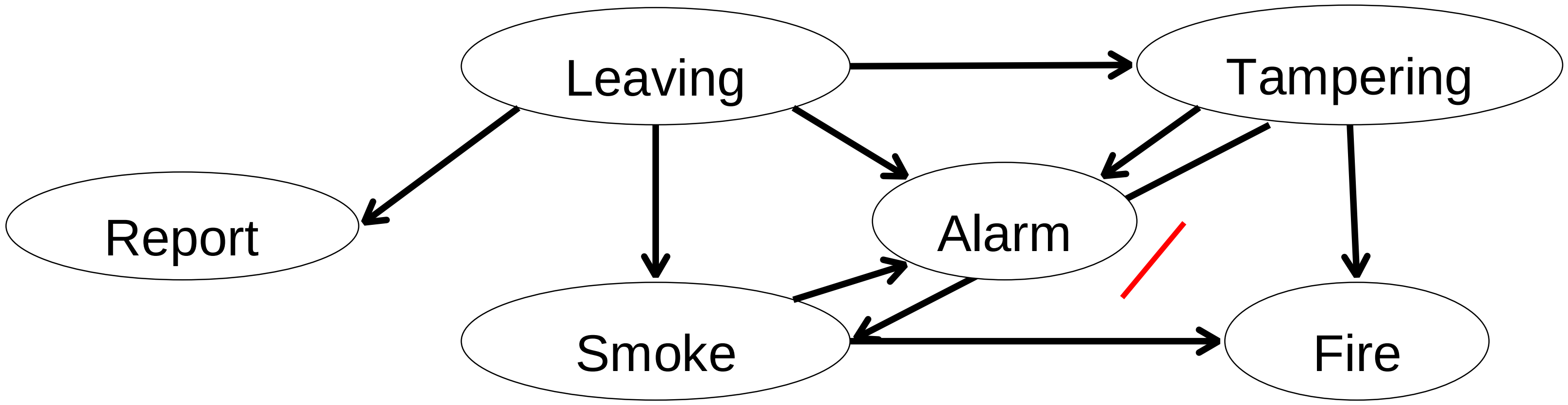


But they are not:
for instance, P(Fire = t| Smoke = f, Tampering = F, Alarm = T) should be higher than P(Fire = t| Smoke = f, Tampering = f),

- How can a network structure be wrong?

Are there wrong network structures?

- If it misses directed edges that are required • E.g. an edge is missing below: Fire is not conditionally independent of Alarm | {Tampering, Smoke}



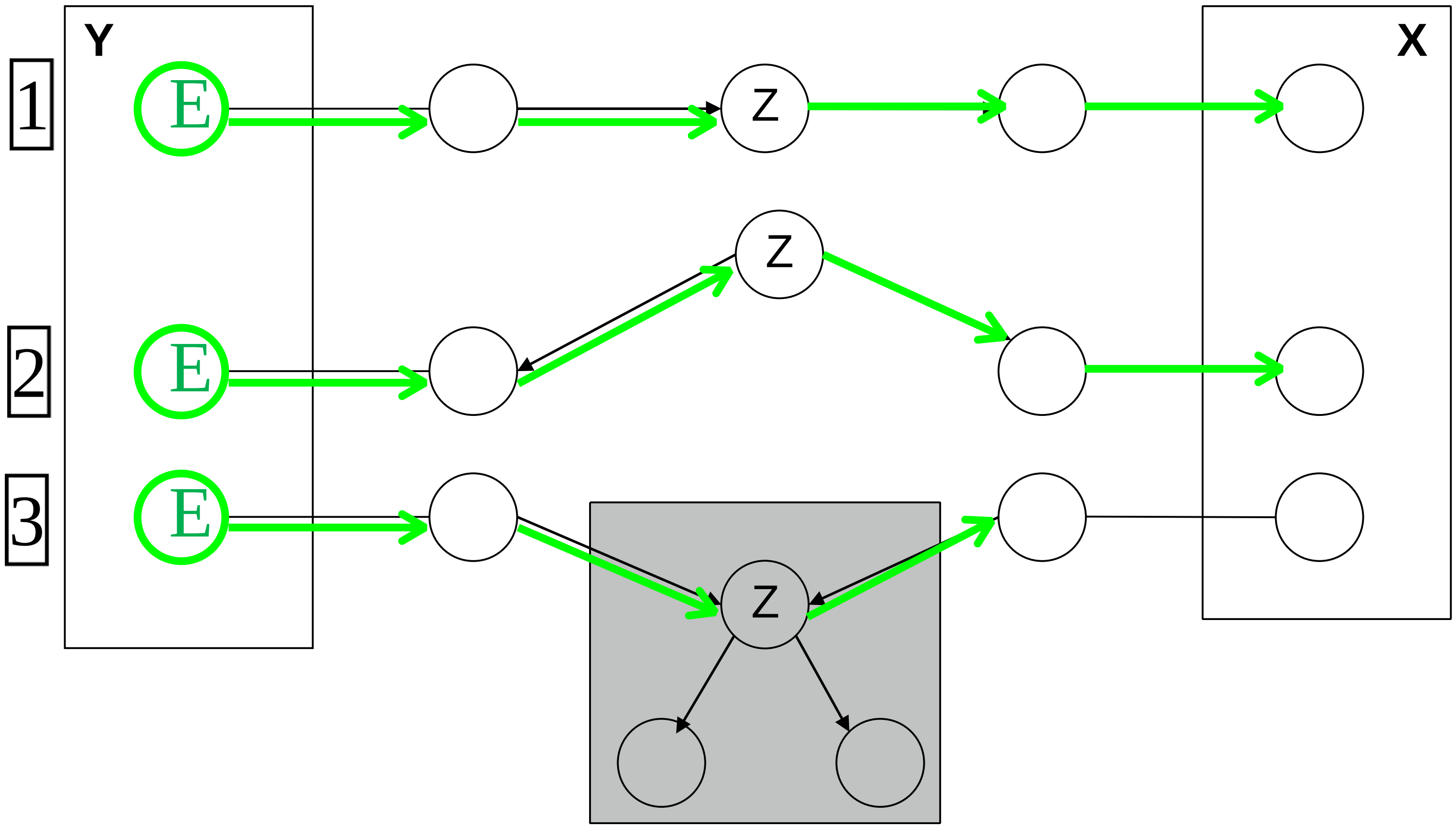
But remember what we said a few slides back. Sometimes we may need to make simplifying assumptions - e.g. assume conditional

Are there wrong network structures?

independence when it does not actually hold - in order to reduce complexity

Summary of Dependencies in a Bayesian Network

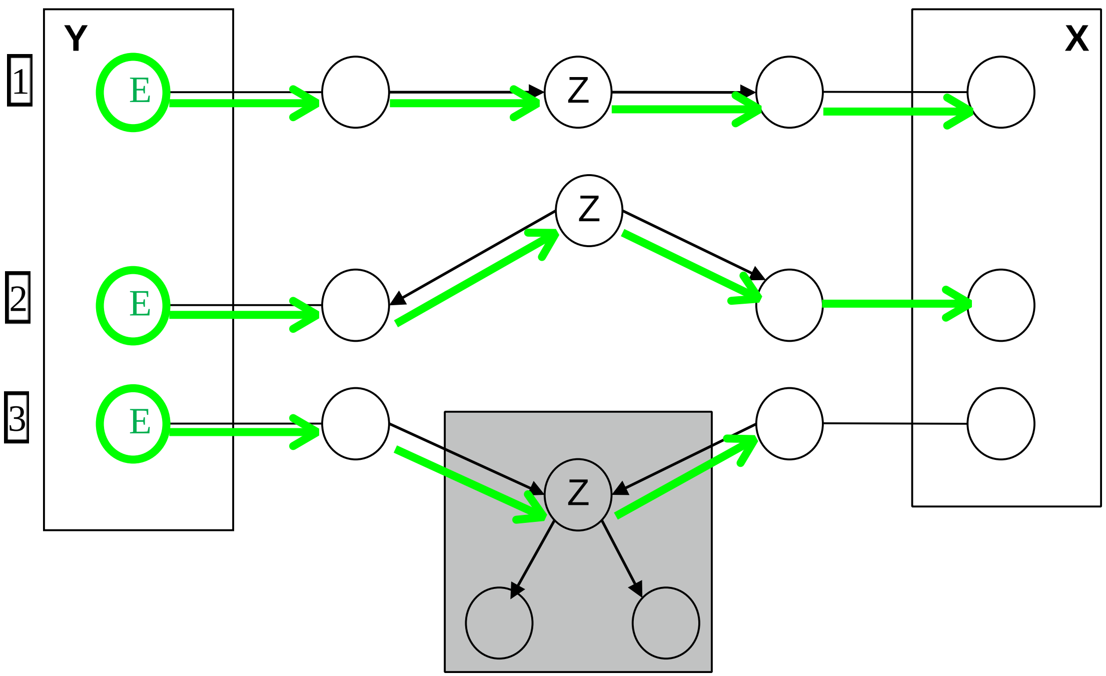
In 1, 2 and 3, X and Y are dependent (grey areas represent existing evidence/observations)

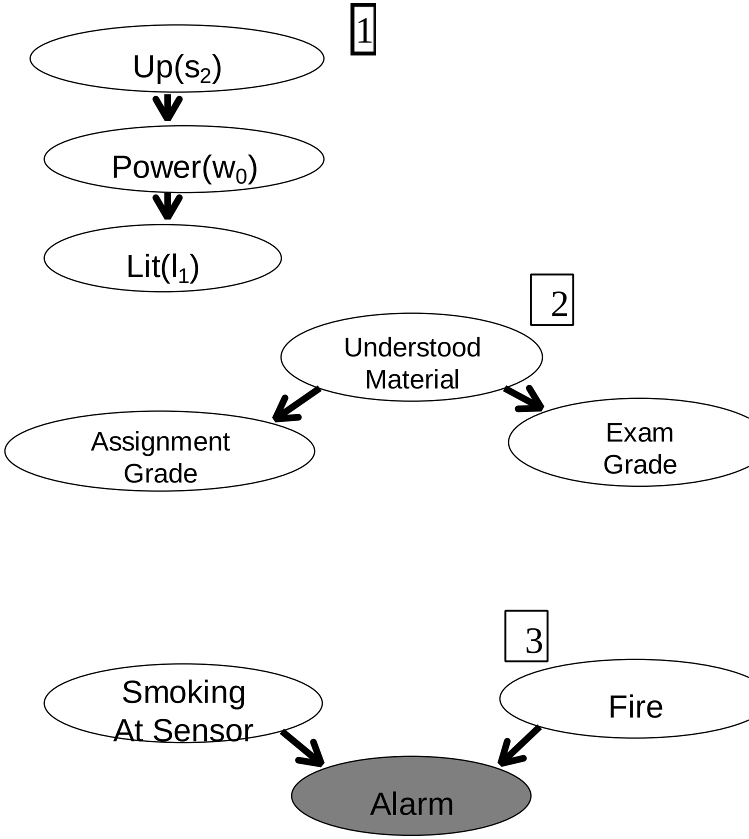


- In 3, X and Y become dependent as soon as there is evidence on Z or on any of its descendants.
- This is because knowledge of one possible cause given evidence of the effect explains away the other cause

Dependencies in a Bayesian Network: summary

In 1, 2 and 3, X and Y are dependent (grey areas represent existing evidence/observations)



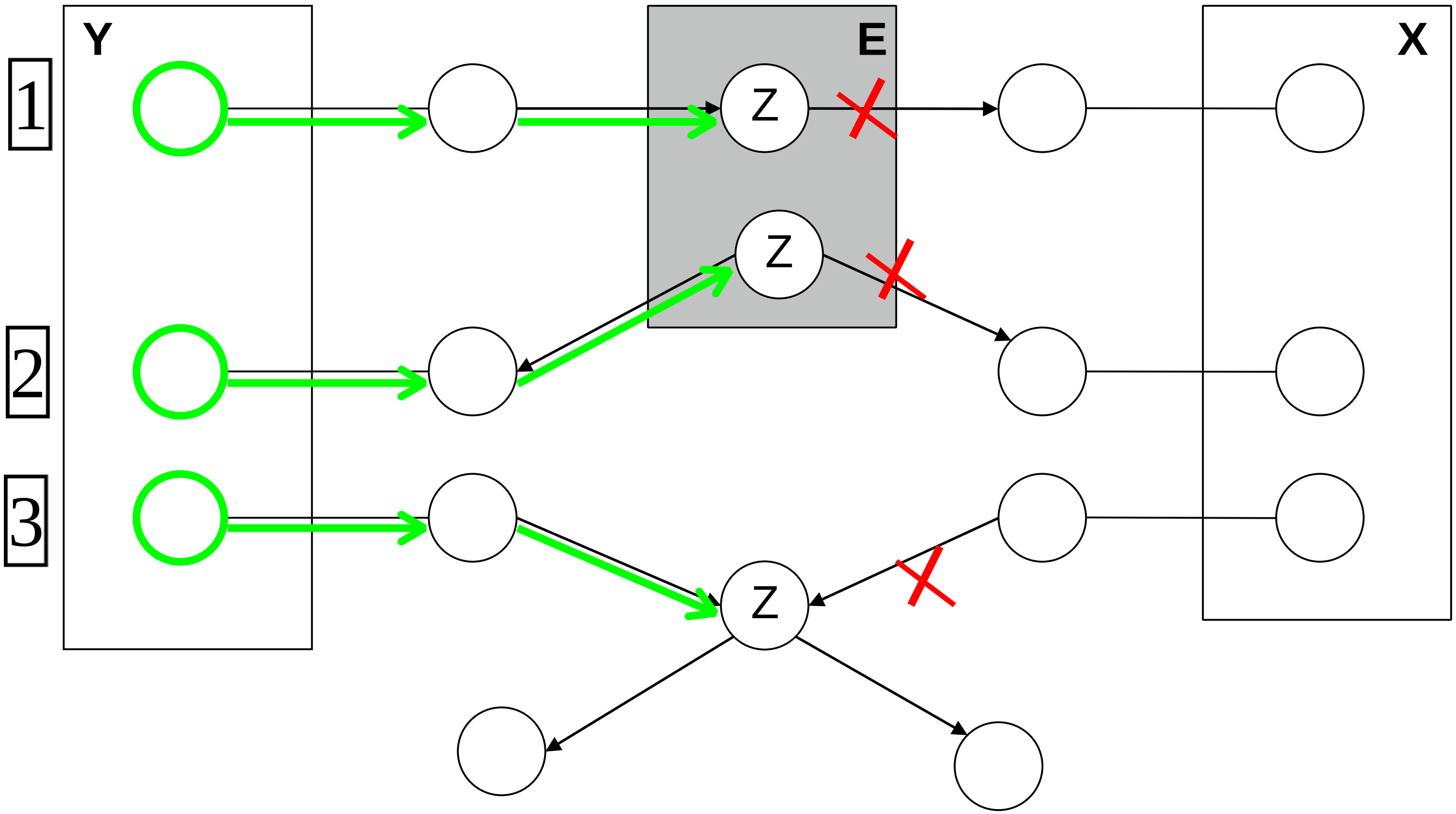


- In 3, X and Y become dependent as soon as there is evidence on Z or on any of its descendants.

- This is because knowledge of one possible cause given evidence of the effect explains away the other cause

Or Conditional Independencies

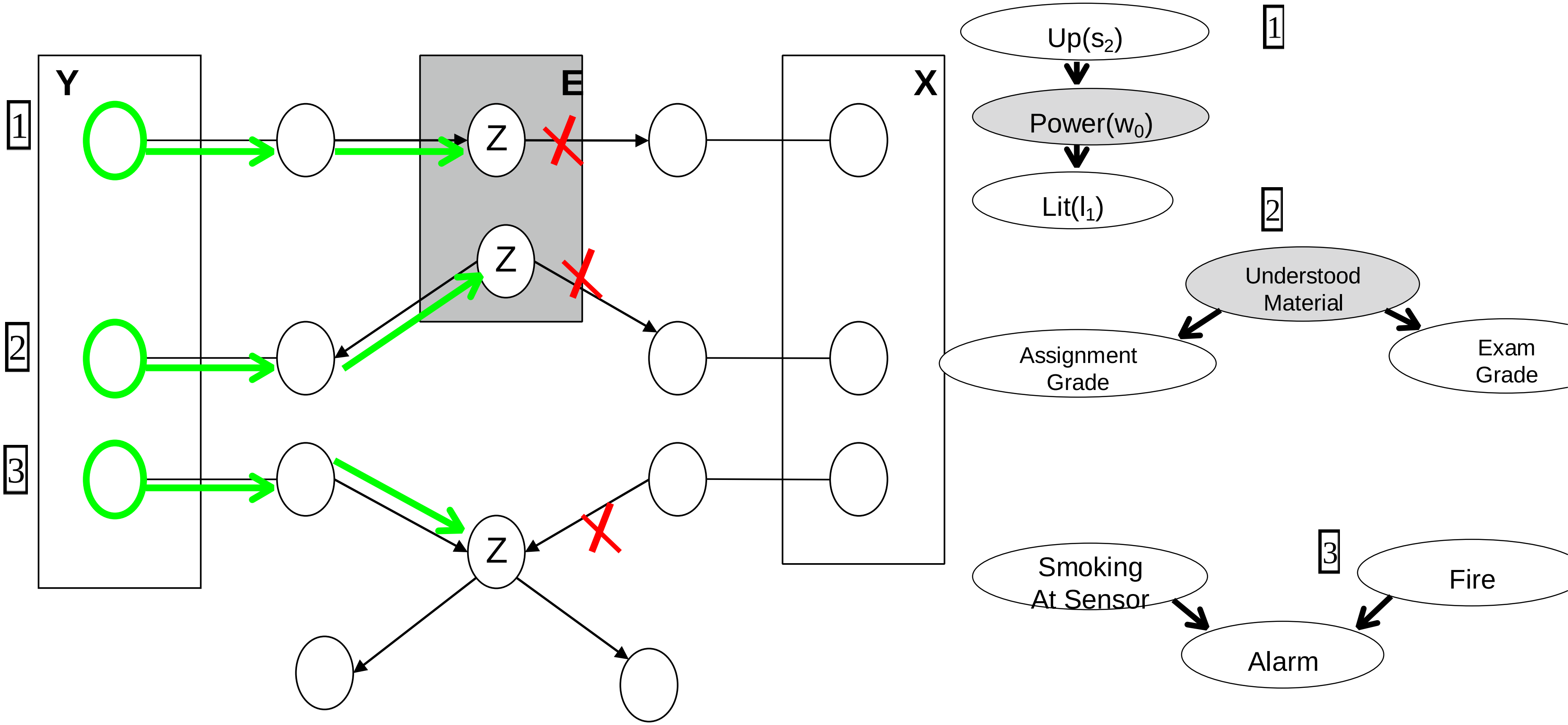
Or, blocking paths for probability propagation. Three ways in which a path between Y to X (or viceversa) can be blocked, given evidence E



- In 3, X and Y are independent if there is no evidence on their common effect (recall fire and tampering in the alarm example

Or Conditional Independencies

Or, blocking paths for probability propagation. Three ways in which a path

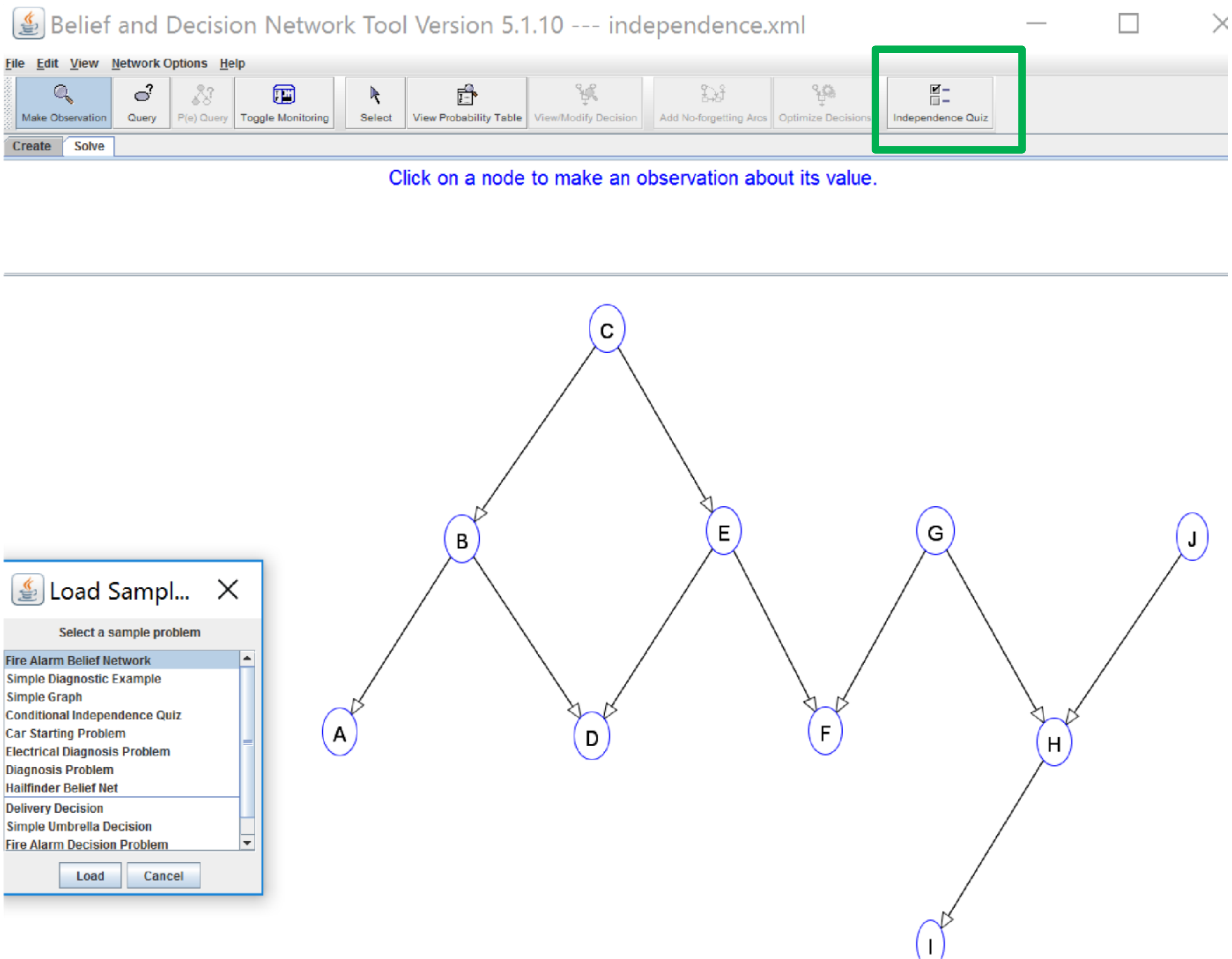


between Y to X (or viceversa) can be blocked, given evidence E

- In 3, X and Y are independent if there is no evidence on their common effect (recall fire and tampering in the alarm example

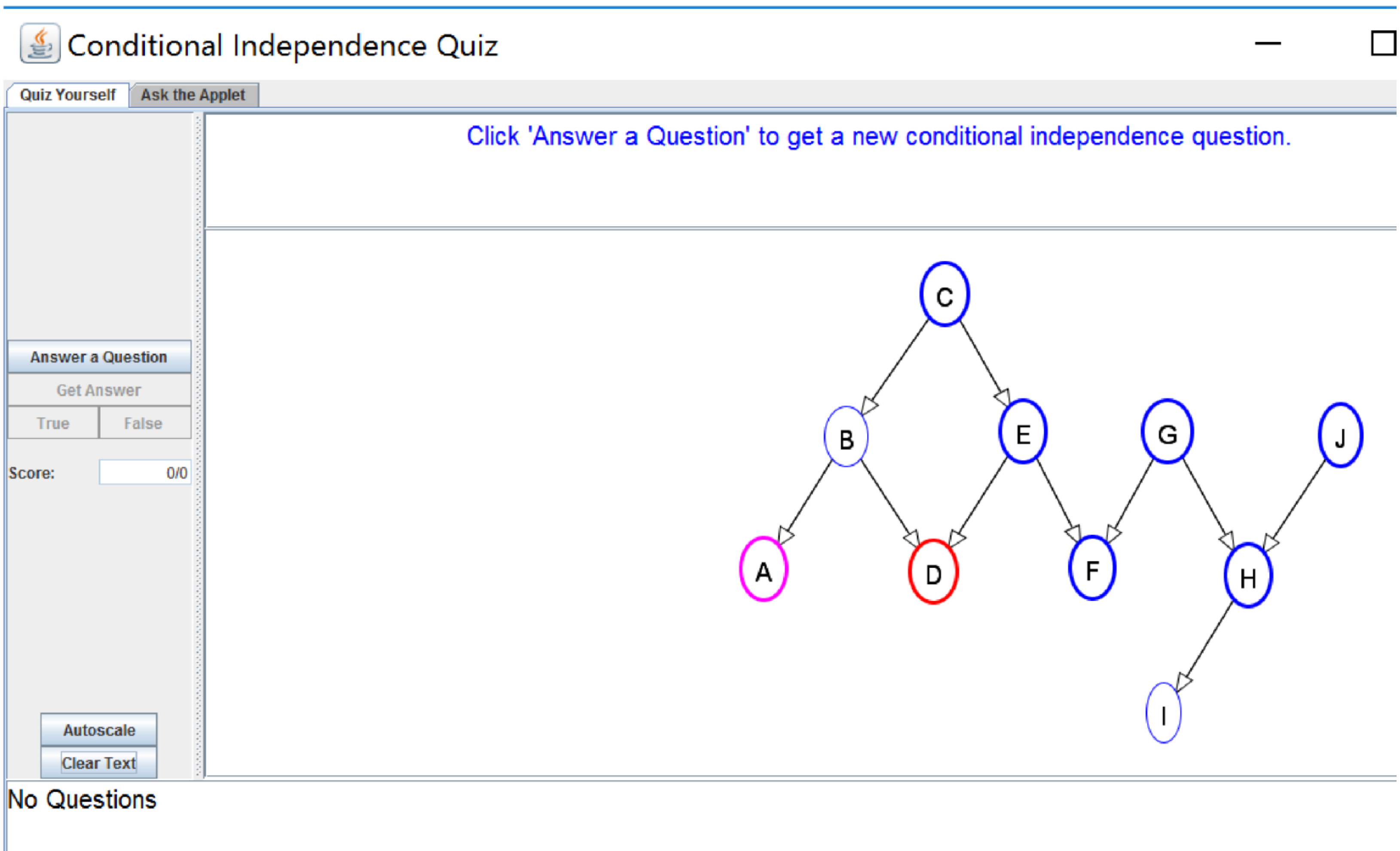
Practice in the AISpace Applet

- Open the Belief and Decision Networks applet
- Load the problem: Conditional Independence Quiz
- Click on Independence Quiz



Practice in the AISpace Applet

- Answer Quizzes in the Conditional Independence Quiz Panel



Learning Goals so Far

- Given a JPD
- Marginalize over specific variables
- Compute distributions over any subset of the variables
- Use inference by enumeration
- to compute joint posterior probability distributions over any subset of variables given evidence
- Define and use marginal and conditional independence
- Build a Bayesian Network for a given domain (structure)
- Specify the necessary conditional probabilities
- Compute the representational savings in terms of number of probabilities required
- Identify dependencies/independencies between nodes in a Bayesian

Network

Now we will see how to do inference in BNETS