Bayesian Networks
Part 3 of 4
Evidence nodes
d-separation & d-connection
Benefits & drawbacks

Evidence nodes

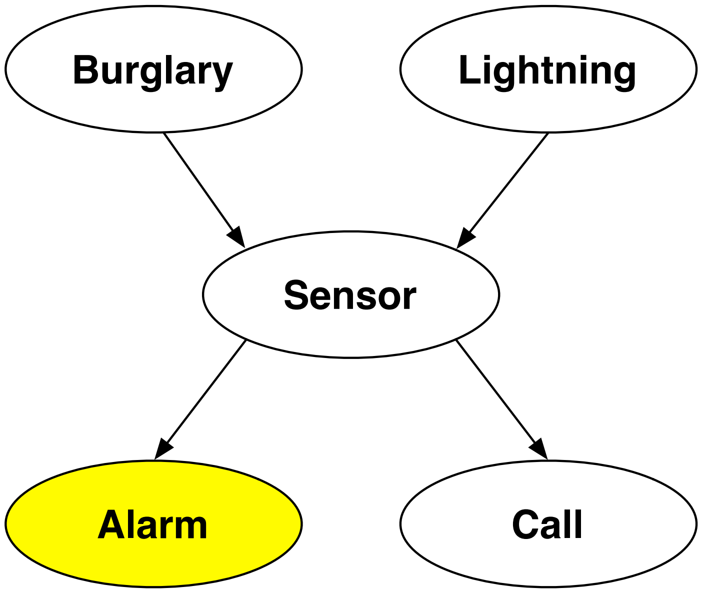
- Given a Bayesian network we can be given the truth or falsity of one or more variables

Evidence nodes – 2

- ◊ Given a Bayesian network we can be given the truth or falsity of one or more variables
  - » **These are called evidence nodes**

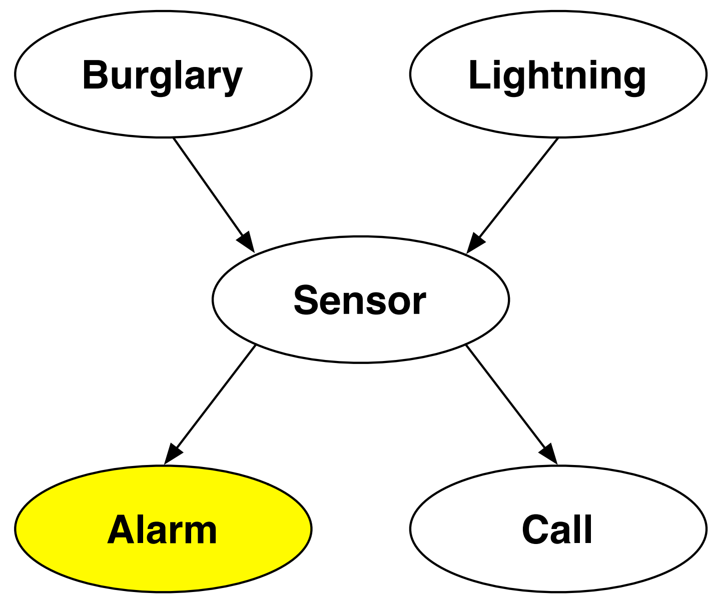
Evidence nodes – 2

- Given a Bayesian network we can be given the truth or falsity of one or more variables
  - **We may learn that an alarm occurred or did not occur**



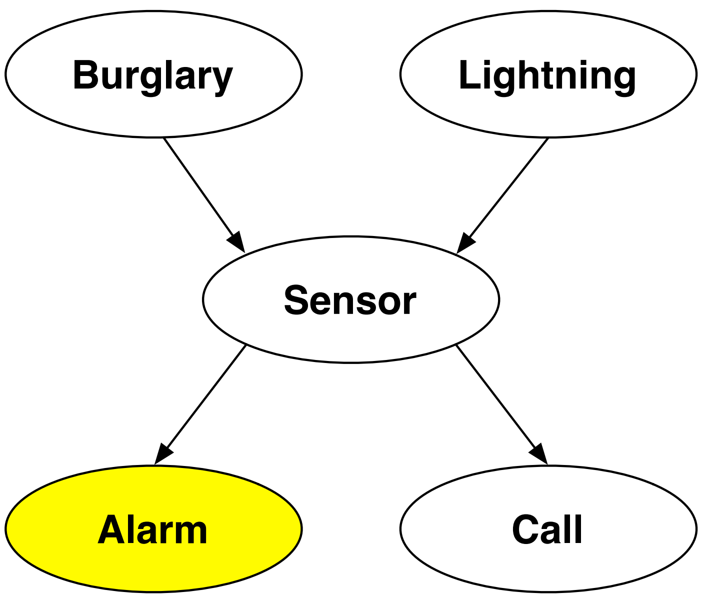
Evidence nodes – 3

- Given a Bayesian network we can be given the truth or falsity of one or more variables
  - We may learn that an alarm occurred or did not occur
    - In which case ‘Alarm’ is an evidence node



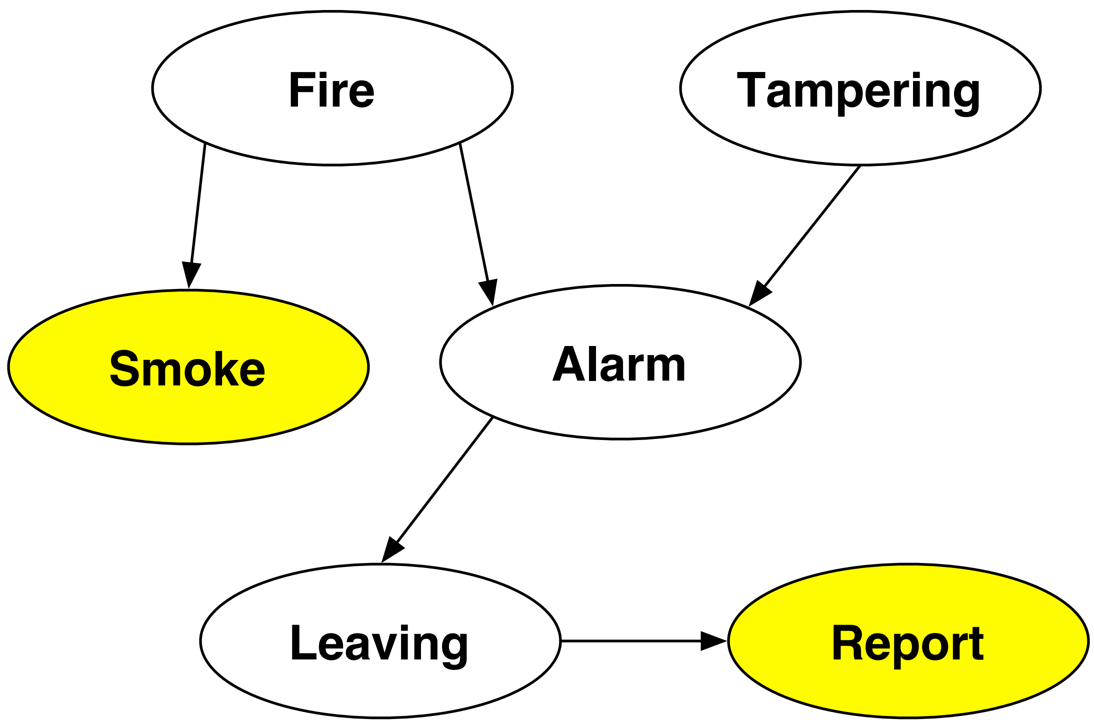
Evidence nodes – 4

- Given a Bayesian network we can be given the truth or falsity of one or more variables
  - » We may learn that an alarm occurred or did not occur
    - > In which case ‘Alarm’ is an evidence node
      - – As a consequence, the probability of the other nodes would change



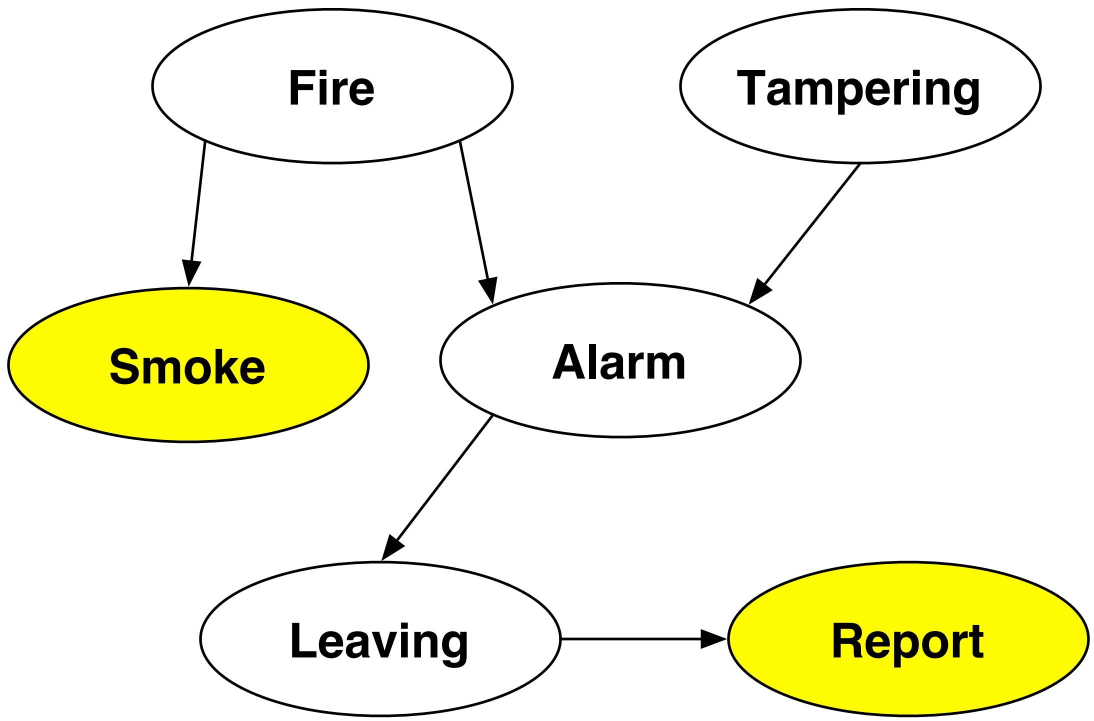
Evidence nodes example 2

◊ Smoke and report could be an evidence set



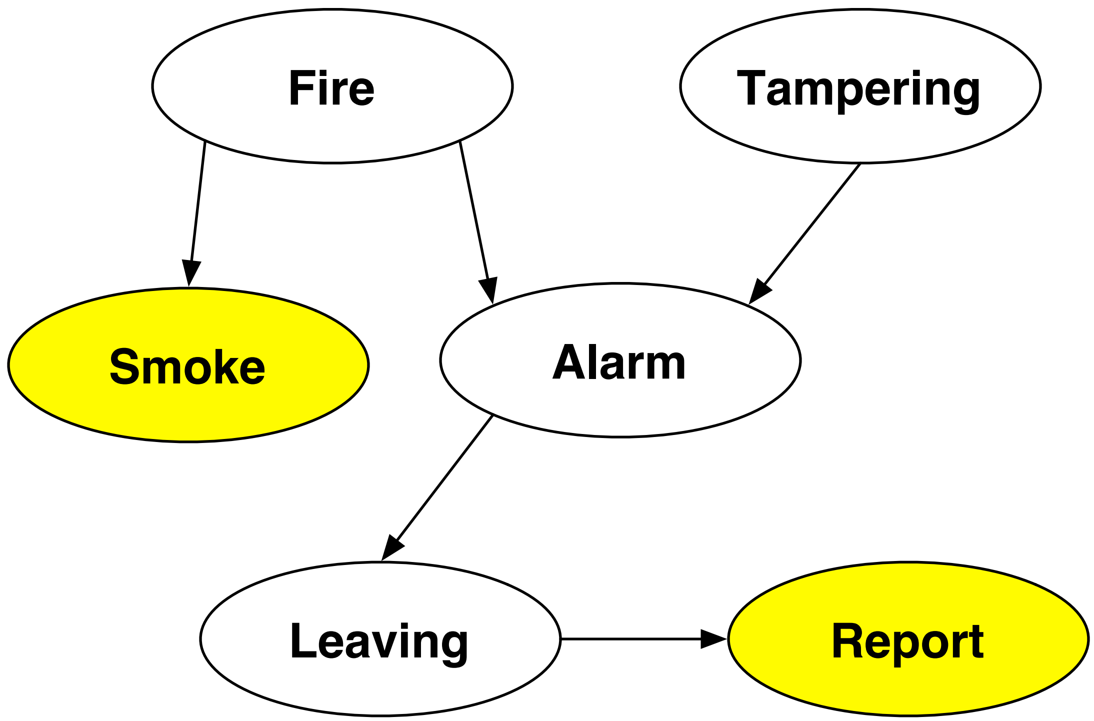
Evidence nodes example 2 – 2

- ◊ Smoke and report could be an evidence set
  - » **You know a report has been submitted and informed that smoke was seen**



Evidence nodes example 2 – 3

- Smoke and report could be an evidence set
  - » **You know a report has been submitted and informed that smoke was seen**
    - > **Increases the probability of a fire and people leaving the building, decreases the probability of tampering**



d-separation & d-connection

◊ Given **evidence nodes** in a Bayesian network and given two nodes $\mathbf{N_j}$ and $\mathbf{N_k}$ in the network

d-separation & d-connection – 2

- Given **evidence nodes** in a Bayesian network and given two nodes N_j and N_k in the network
  - **Are the probabilities of the variables dependent or independent?**
    - **Why do we want to know?**

d-separation & d-connection – Why

- Given **evidence nodes** in a Bayesian network and given two nodes N_j and N_k in the network
  - » **Are the probabilities of the variables dependent or independent?**

To simplify equations, simplify computation

d-separation & d-connection – Why – 2

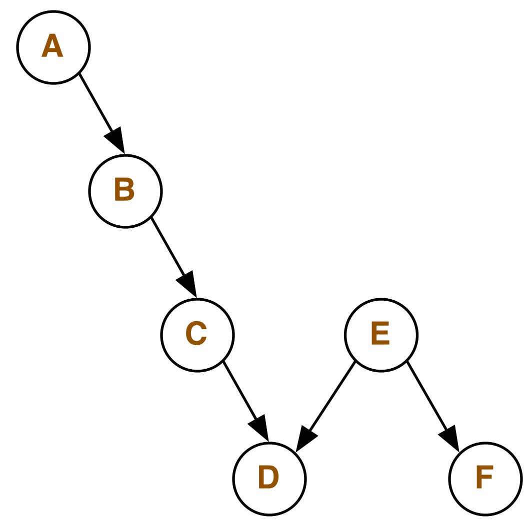
◊ Given **evidence nodes** in a Bayesian network and given two nodes N_j and N_k in the network

» **Are the probabilities of the variables dependent or independent?**

To simplify equations, simplify computation.
Have to know when simplification can be done.

» **P(C | A ^ B) → P(C | B)**

» **P(C | B ^ D ^ F)   no simplification**



d-separation & d-connection – 3

◊ Given **evidence nodes** in a Bayesian network and given two nodes N_j and N_k in the network

» **Are the probabilities of variables dependent or independent?**

◊ We speak of

» **d-separation of the variables**

d-separation & d-connection – 4

◊ Given **evidence nodes** in a Bayesian network and given two nodes N_j and N_k in the network

» **Are the probabilities of variables dependent or independent (separate)?**

◊ We speak of

» **d-separation of the variables**

> **direction-dependent separation**

d-separation & d-connection – 5

- Given **evidence nodes** in a Bayesian network and given two nodes **N_j** and **N_k** in the network
  - » **Are the probabilities of variables dependent or independent (separate)?**
- We speak of
  - » **d-separation of the variables**
    - > **direction-dependent separation**
  - » **Variables that are not d-separated are said to be d-connected**

d-separation definition

◊ Given an evidence set **E**, Nodes $\mathbf{N_j}$ and $\mathbf{N_k}$ are said to be conditionally independent if **E** d-separates $\mathbf{N_j}$ and $\mathbf{N_k}$

d-separation definition – 2

◊ Given and evidence set **E**, Nodes $\mathbf{N_j}$ and $\mathbf{N_k}$ are said to be conditionally independent if **E** d-separates $\mathbf{N_j}$ and $\mathbf{N_k}$

◊ **E** d-separates $\mathbf{N_j}$ and $\mathbf{N_k}$

» **If all undirected paths ($\mathbf{N_j}$, $\mathbf{N_k}$) are blocked by E**

d-separation definition – 3

- Given and evidence set **E**, Nodes **N_j** and **N_k** are said to be conditionally independent if **E** d-separates **N_j** and **N_k**
- **E** d-separates **N_j** and **N_k** if all **undirected** paths (**N_j , N_k**) are **blocked** by **E**
- If **E** d-separates **N_j** and **N_k** then
  - **N_j and N_k are conditionally independent**

d-separation definition – 4

- Given and evidence set **E**, Nodes N_j and N_k are said to be conditionally independent if **E** d-separates N_j and N_k
- **E** d-separates N_j and N_k if all **undirected** paths (N_j , N_k) are **blocked** by **E**
- If **E** d-separates N_j and N_k then
  - » **N_j and N_k are conditionally independent**
- We write $I (N_j , N_k \mid E)$ – (I)ndependent
  - » $p (N_j , N_k \mid E) = p (N_j \mid E) * p (N_k \mid E)$

Evidence nodes blocking a path

◊ A path between N_j and N_k is **blocked** by nodes **E**

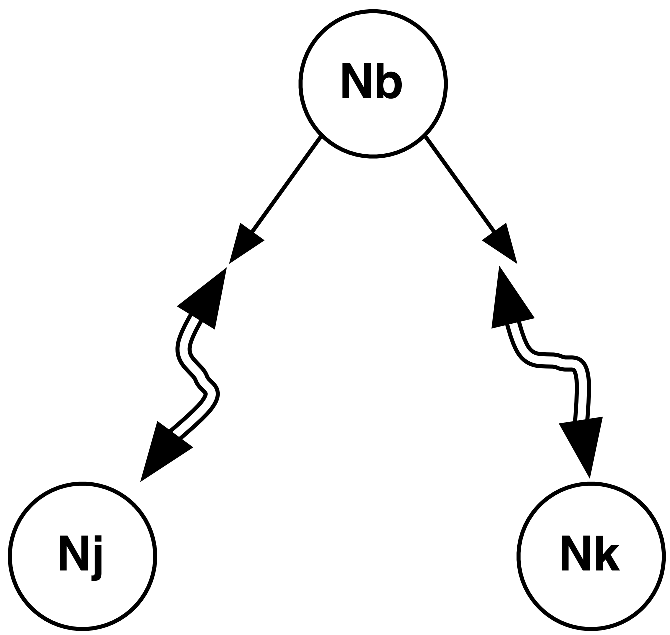
» **If one of the following 3 conditions holds**

> $N_b \in E$ **and both edges on the path lead out of** N_b

Common cause blocking

◊ N_b is a common cause

> $N_b \in E$ **and both edge on the path lead out of N_b**



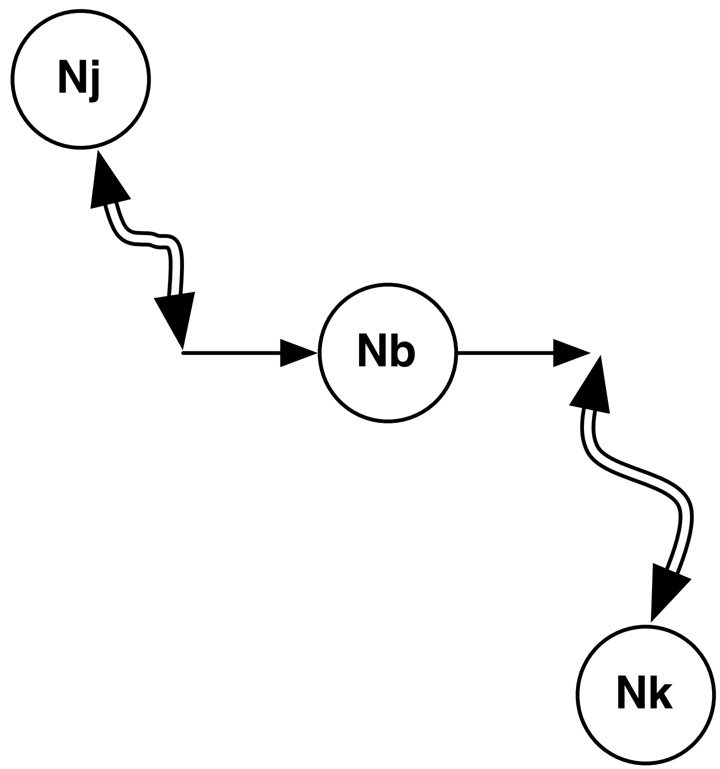
Evidence nodes blocking a path – 2

- A path between N_j and N_k is **blocked** by nodes **E**
  - **If one of the following 3 conditions holds**
    - > $N_b \in E$ **and both edges on the path lead out of** N_b
    - > $N_b \in E$ **and one edge on the path leads into** N_b **and one edge leads out of** N_b

More direct cause

◊ N_b is a more direct (closer) cause of N_k than N_j

> $N_b \in E$ **and one edge on the path leads into N_b and one edge leads out of N_b**



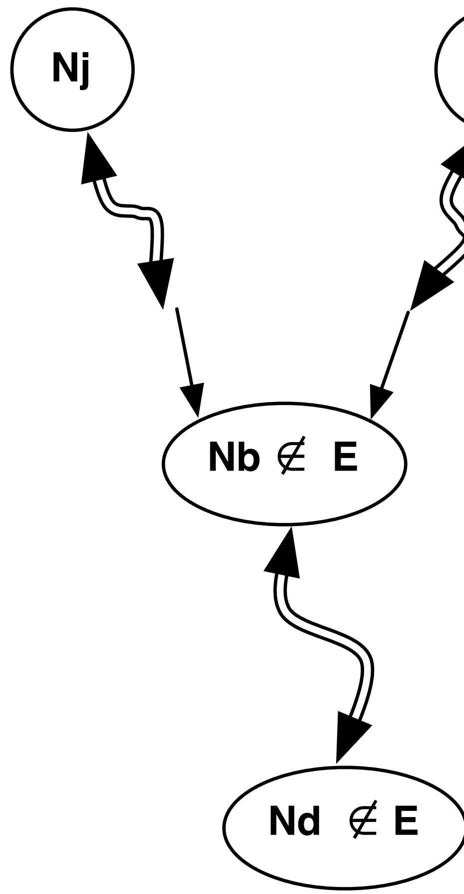
Evidence nodes blocking a path – 3

- A path between N_j and N_k is **blocked** by nodes **E**
  - **If one of the following 3 conditions holds**
    - **$N_b \in E$ and both edges on the path lead out of N_b**
    - **$N_b \in E$ and one edge on the path leads into N_b and one edge leads out of N_b**
    - **Neither N_b nor any descendent of N_b is in E and both edges on the path lead into N_b**

Common consequence

◊ N_b is a common consequence of

> **Neither N_b nor any descendent of N_b is in E and both edges on the path lead into N_b**



Benefits

◊ Based on sound mathematics of probability theory

Benefits – 2

◊ Based on sound mathematics of probability theory

◊ Can reason in both the forward and backward directions.

Benefits – 3

◊ Based on sound mathematics of probability theory

◊ Can reason in both the forward and backward directions.

» **Given causes can compute probability of consequences**

Benefits – 4

- ◊ Based on sound mathematics of probability theory
- ◊ Can reason in both the forward and backward directions.
  - » **Given causes can compute probability of consequences**
  - » **Given consequences can estimate probability of different causes**

Benefits – 5

◊ Casualty network structure eliminates the need to compute probabilities for all combinations of all variables

Benefits – 6

- ◊ Casualty network structure eliminates the need to compute probabilities for all combinations of all variables
  - » **Example**
    - > **Burglar network has 5 nodes that require 10 probability estimates over all the tables**

Benefits – 7

- ◊ Casualty network structure eliminates the need to compute probabilities for all combinations of all variables
  - » **Example**
    - > **Burglar network has 5 nodes that require 10 probability estimates over all the tables**
    - > **All combinations of 5 variables would require 2^5 – 1 = 31 probability estimates**

Benefits – 8

- ◊ Casualty network structure eliminates the need to compute probabilities for all combinations of all variables
  - » Example
    - > Burglar network has 5 nodes that require 10 probability estimates over all the tables
    - > All combinations of 5 variables would require 2^5 – 1 = 31 probability estimates
      - – Probabilities must add to 1, so last number can be computed

Benefits – 9

- ◊ Casualty network structure eliminates the need to compute probabilities for all combinations of all variables
  - » Example
    - > Burglar network has 5 nodes that require 10 probability estimates over all the tables
    - > All combinations of 5 variables would require 2^5 – 1 = 31 probability estimates
      - – Probabilities must add to 1, so last number can be computed
    - > Structure is equivalent to 21 numbers

Benefits – 10

- Casualty network structure eliminates the need to compute probabilities for all combinations of all variables
  - **Example**
    - **Burglar network has 5 nodes that require 10 probability estimates over all the tables**
    - **All combinations of 5 variables would require 2^5 – 1 = 31 probability estimates**
      - **Probabilities must add to 1, so last number can be computed**
    - **Structure is equivalent to 21 numbers**
  - **Much larger savings as the network grows**

Benefits – 11

- ◊ Casualty network structure eliminates the need to compute probabilities for all combinations of all variables
  - » **Example**
    - > **Burglar network has 5 nodes that require 10 probability estimates over all the tables**
    - > **All combinations of 5 variables would require 2^5 – 1 = 31 probability estimates**
      - – **Probabilities must add to 1, so last number can be computed**
    - > **Structure is equivalent to 21 numbers**
  - » **Much larger savings as the network grows**
    - > **Can handle significantly larger models**

Drawbacks

◊ Complexity of reasoning grows exponentially with the number of nodes

Drawbacks – 2

◊ Complexity of reasoning grows exponentially with the number of nodes

◊ Propagation is complex

Drawbacks – 3

- Complexity of reasoning grows exponentially with the number of nodes
- Propagation is complex
  - **Large networks**

Drawbacks – 4

◊ Complexity of reasoning grows exponentially with the number of nodes

◊ Propagation is complex

  » **Large networks**

  » **No single algorithm for all networks**

Drawbacks – 5

- ◊ Complexity of reasoning grows exponentially with the number of nodes
- ◊ Propagation is complex
  - » **Large networks**
  - » **No single algorithm for all networks**
- ◊ Mitigation
  - » **Modern algorithms use variable elimination to carry out modular calculations on parts of the model, which are then combined**

Drawbacks – 6

◊ Complexity of reasoning grows exponentially with the number of nodes

◊ Propagation is complex

  » **Large networks**

  » **No single algorithm for all networks**

◊ Mitigation

  » **Modern algorithms use variable elimination to carry out modular calculations on parts of the model, which are then combined**

    > **Rather than working on the whole model as a single entity**

Drawbacks – 7

- ◊ Entering probability tables for large models

Drawbacks – 8

- ◊ Entering probability tables for large models
  - » **Too many individual numbers**

Drawbacks – 9

◊ Entering probability tables for large models
  » **Too many individual numbers**
  » **Too complex if there are many parents**

Drawbacks – 10

◊ Entering probability tables for large models

» **Too many individual numbers**

» **Too complex if there are many parents**

◊ Mitigation

» **Use expressions to compute probability values**