Bayesian Networks Representation

Dave Essary

Overview

- Distributions, graphs, and independences
- I-MAPs
- d-separation (d-connection)
- I-equivalence
- P-MAP and equivalence classes

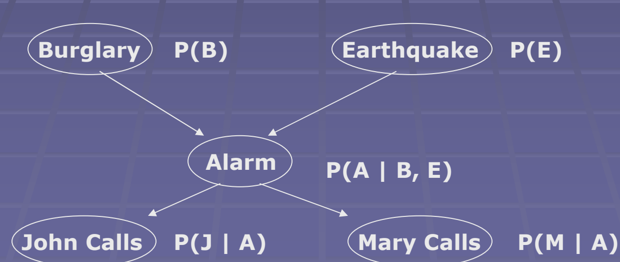
Independences

- Distribution: a set of probabilities of a random variables representing our view of the world
- We use graphs to help us visualize the dependences and independences between these variables.
- Independences allow us to factor out variables
  - Ex: $P(X, Y) = P(X) * P(Y)$ if $(X \perp Y)$
  - Local probability tables become smaller
- POINT:

  Independences = Smaller Tables = GOOD!

Independences in BBN (2)

- Graph: DAG where nodes represent random variables and edges (missing edges) represent dependences (independences)
- A BBN is a graph G and a set of local conditional probability distributions (CPDs)



Independences in BBN (3)

- Local Markov assumption:
  - For each variable X_i, given its parents, $Pa[X_i]$, all non-descendants of X_i are independent of X_i.
  - Notation: $(X_i \perp NonDescendants[X_i] \mid Pa[X_i])$
- Consequence: only need local probability distributions to represent the full joint
  - We can recover the full joint distribution whenever we want

Independences in BBN (4)

- A specific graph (structure) defines a class of probability distributions that obey certain independence assumptions

Overview

- Distributions, graphs, and independences
- I-MAPs
- d-separation (d-connection)
- I-equivalence
- P-MAP and equivalence classes

Independence Assertions

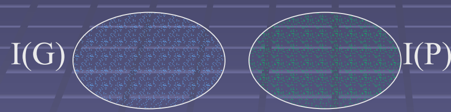
- Given distribution P. Let I(P) be the set of independence relations of the form

  $(X \perp Y \mid Z)$

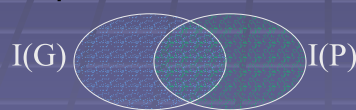
  that hold in P.
- Similarly, given graph G, let I(G) be the set of independence assertions that are expressed in graph G.

Independence Assertions (2)

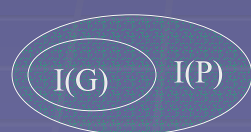
- I(G) and I(P) may be disjoint…



- Overlap…



- Or be subsets of each other…



I-MAPs

- BN structure G is an I-MAP (independence mapping) of P if $I(G) \subseteq I(P)$
- Note: complete graphs are always I-MAPs
  - Since all nodes are connected, there are no independences
  - $I(G_{Complete}) = \varnothing \subseteq I(P)$

Factorization

- Distribution P over X is said to factorize according to G if P can be expressed as a product of each variable's probability, given its parents' values…

  $P(X_1 \ldots X_n) = \prod P(X_i \mid Pa[X_i])$

  - This is the chain rule for BNs
  - G is an I-map for P if and only if P factorizes according to G.
- A BN is a pair (G, P) where P factorizes according to G, and where P is specified as a set of CPDs associated with G's nodes.

Overview

- Distributions, graphs, and independences
- I-MAPs
- d-separation
- I-equivalence
- P-MAP and equivalence classes

D-separation

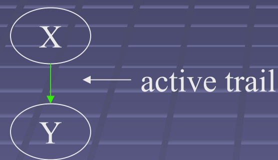
- A graphical criterion used to identify independences (marginal or conditional) that hold in the BBN graph

Active Trails

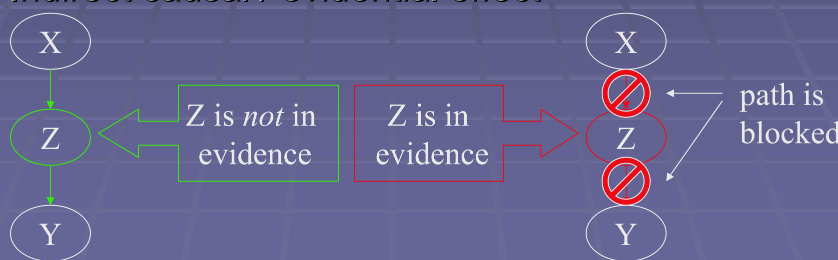
- Active trails can occur along a few different paths between nodes…
  - Direct connection
    - $X \rightarrow Y$ or $X \leftarrow Y$
  - Indirect causal / evidential effect
    - $X \rightarrow Z \rightarrow Y$ or $X \leftarrow Z \leftarrow Y$ without Z
  - Common cause
    - $X \leftarrow Z \rightarrow Y$ without Z
  - Common effect (also called a v-structure)
    - $X \rightarrow Z \leftarrow Y$ given Z or any of its descendants

Active Trails (2)

- Direct connection

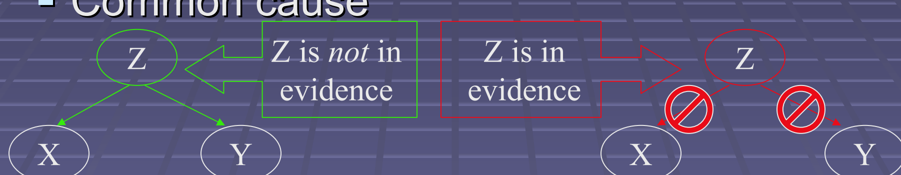


- Indirect causal / evidential effect

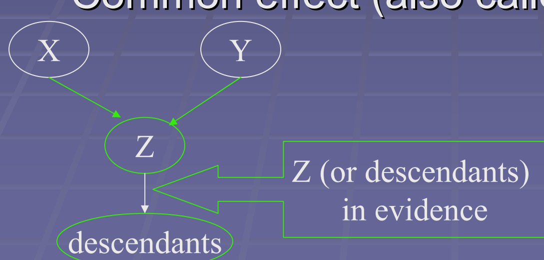


Active Trails (3)

- Common cause

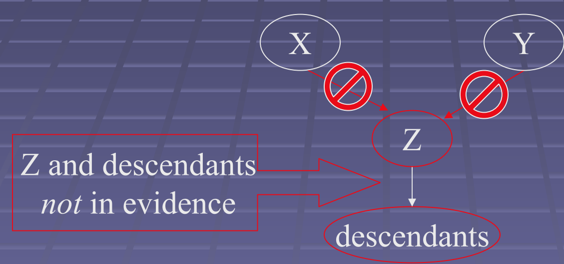


- Common effect (also called a v-structure)



Active Trails (4)

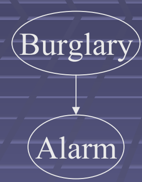
- Common effect (cont.)



d-separation

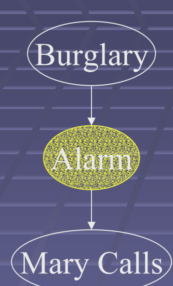
- Two nodes in a BN are d-separated if there is no active trail between them
- If two nodes X and Y in G are d-separated (given Z), then they are conditionally independent in all distributions that factorize over G
- If two nodes X and Y in G are *not* d-separated (given Z), then X and Y are dependent in *some* distribution P that factorizes over G

Direct Connection



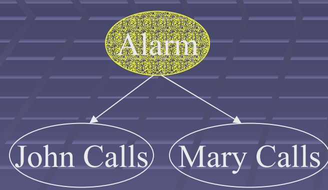
Obviously, Alarm will always depend on the value of Burglary

Indirect Effect



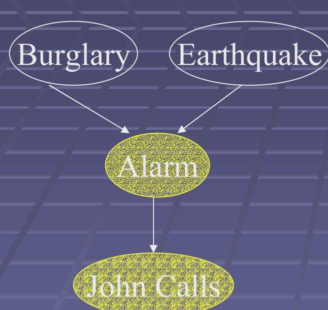
- Q: Assume we know that the alarm has gone off. Mary then calls. Does this affect our belief in a burglary?
- A: No, it shouldn't. Mary calling gives us no new information. Hence, Burglary and Mary Calls are conditionally independent.

Common Cause



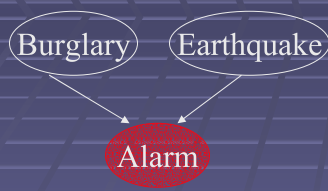
- Q: Assume we know that the alarm has gone off. John then calls. Does this affect our belief in Mary calling?
- A: No, it shouldn't. Since we already know the alarm has gone off, our belief in Mary calling is based strictly on whether we believe Mary is paying attention. Hence, Mary calling and John calling are independent.

Common Effect (v-structure)



- Q: Assume we know that the alarm has gone off (or we suspect because a friend called us). This obviously has an effect on our belief in burglary. We then find out that an earthquake has occurred at our house. Does this affect our belief in a burglary occurring?
- A: Yes, it does. The alarm going off increases our belief in a burglary. BUT, once we know the earthquake has occurred, we then "explain away" the alarm, and our belief in a burglary drops down again. Hence, Burglary and Earthquake are dependent upon each other.

Common Effect (v-structure)



- Q: Now assume we know that we do *not* know whether the alarm has gone off or not. We then find out that an earthquake has occurred at our house. Does this affect our belief in a burglary occurring?
- A: No, it shouldn't. A burglary and an earthquake should be independent events.

Testing for d-separation

- How can we test for d-separation?
- Enumerate over all paths.
  - Bad idea… number of paths is exponential in size of the graph.
- Better strategy: Two step sweep
  - Step 1: Begin at leaves. Traverse graph bottom up, marking all nodes in evidence E or that have descendents in E.
  - Step 2: Begin at source node X. Traverse graph in BFS fashion, stopping at a node N if
    - N is the "middle" of a v-structure and is unmarked
    - N is not the "middle" of a v-structure and is in E
- This strategy has a running time linear in the size of the graph

Overview

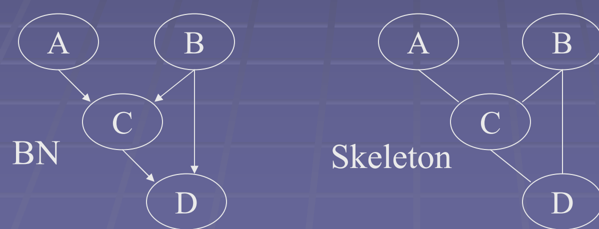
- Distributions, graphs, and independences
- I-MAPs
- d-separation (d-connection)
- I-equivalence
- P-MAP and equivalence classes

I-equivalence

- Two BN graphs G_1 and G_2 over X are I-equivalent if $I(G_1) = I(G_2)$
- The set of all graphs over X is partitioned into a set of mutually exclusive and exhaustive I-equivalence classes.

I-equivalence (2)

- How do we test I-equivalence?
- Skeleton
  - Undirected version of a BN G that has an edge {X,Y} for each directed edge (X,Y) in G.



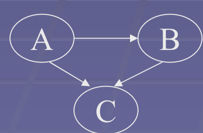
I-equivalence (3)

- If G_1 and G_2 have the same skeleton and v-structures, then they are I-equivalent
- Reverse is not necessarily true!
  - All complete graphs are I-equivalent, but they do not necessarily have the same skeleton or v-structures.

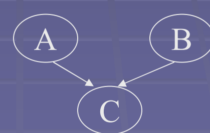
I-equivalence (4)

- Immorality
  - A v-structure $X \rightarrow Z \leftarrow Y$ is an immorality if there is no direct edge between X and Y.
  - G_1 and G_2 have the same skeleton and immoralities if and only if they are I-equivalent

v-structure



v-structure and immorality



Overview

- Distributions, graphs, and independences
- I-MAPs
- d-separation (d-connection)
- I-equivalence
- P-MAP and equivalence classes

P-MAP

- We want to go from a distribution to a graph. How can we do this?
- Minimal I-MAP
  - I-MAP G for P is a minimal I-MAP if the removal of a single edge results in G no longer being an I-MAP
  - Not necessarily a good candidate for capturing independences of a distribution

P-MAP (2)

- P-MAP (or perfect map) G for distribution P has the property I(G) = I(P)
- Not all distributions have a P-MAP
  - Ex: X xor Y xor Z
    - Given no evidence, all variables are independent of each other
    - Given any single variable, the other two become dependent on each other
- Finding a P-MAP, or, more useful, an equivalence class of P-MAPs, is no more difficult than finding a skeleton and the immoralities of a P-MAP.