Reasoning with Graphical Models

Slides Set 3:

Rina Dechter

Reading:
Darwiche chapter 4
Pearl: chapter 3

Outline

- DAGS, Markov(G), Bayesian networks
- Graphoids: axioms of for inferring conditional independence (CI)
- D-separation: Inferring CIs in graphs
  - I-maps, D-maps, perfect maps
  - Markov boundary and blanket
  - Markov networks

Properties of Probabilistic independence

THEOREM 1: Let X, Y, and Z be three disjoint subsets of variables from U. If $I(X, Z, Y)$ stands for the relation "X is independent of Y, given Z" in some probabilistic model P, then I must satisfy the following four independent conditions:

- Symmetry:
  - **I(X,Z,Y) → I(Y,Z,X)**
- Decomposition:
  - **I(X,Z,YW)→ I(X,Z,Y) and I(X,Z,W)**
- Weak union:
  - **I(X,Z,YW)→I(X,ZW,Y)**
- Contraction:
  - **I(X,Z,Y) and I(X,ZY,W)→I(X,Z,YW)**
- Intersection:
  - **I(X,ZY,W) and I(X,ZW,Y) → I(X,Z,YW)**

Graphoid axioms:
Symmetry, decomposition
Weak union and contraction

Positive graphoid:
+intersection

In Pearl: the 5 axioms
are called Graphids,
the 4, semi-graphois

Intersection

Holds only for strictly positive distributions

$I_{\Pr}(\mathbf{X}, \mathbf{Z} \cup \mathbf{W}, \mathbf{Y})$ and $I_{\Pr}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$ only if $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

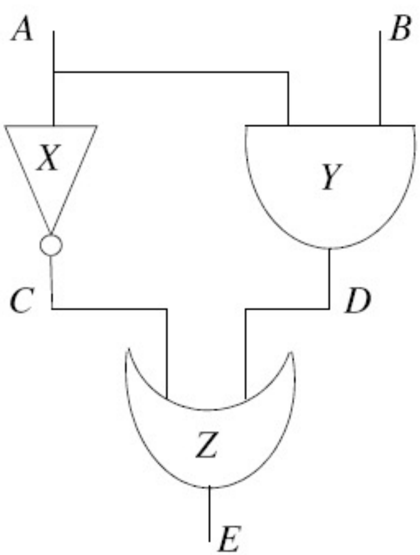
If information **w** is irrelevant given **y**, and **y** is irrelevant given **w**, then combined information **yw** is irrelevant to start with.

Intersection

Holds only for strictly positive distributions

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z} \cup \mathbf{W}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$ only if $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

If information **w** is irrelevant given **y**, and **y** is irrelevant given **w**, then combined information **yw** is irrelevant to start with.



- If we know the input A of inverter X, its output C becomes irrelevant to our belief in the circuit output E.
- If we know the output C of inverter X, its input A becomes irrelevant to this belief.
- Yet, variables A and C are not irrelevant to our belief in the circuit output E.

Properties of Probabilistic independence

THEOREM 1: Let X, Y, and Z be three disjoint subsets of variables from U. If $I(X, Z, Y)$ stands for the relation "X is independent of Y, given Z" in some probabilistic model P, then I must satisfy the following four independent conditions:

- Symmetry:
  - I(X,Z,Y) → I(Y,Z,X)
- Decomposition:
  - I(X,Z,YW)→ I(X,Z,Y) and I(X,Z,W)
- Weak union:
  - I(X,Z,YW)→I(X,ZW,Y)
- Contraction:
  - I(X,Z,Y) and I(X,ZY,W)→I(X,Z,YW)
- Intersection:
  - I(X,ZY,W) and I(X,ZW,Y) → I(X,Z,YW)

Graphoid axioms:
Symmetry, decomposition
Weak union and contraction

Positive graphoid:
+intersection

In Pearl: the 5 axioms
are called Graphids,
the 4, semi-graphois

Outline

- DAGS, Markov(G), Bayesian networks
- Graphoids: axioms of for inferring conditional independence (CI)
- D-separation: Inferring CIs in graphs
  - I-maps, D-maps, perfect maps
  - Markov boundary and blanket
  - Markov networks

Outline

- Bayesian Networks, DAGS, Markov(G)
- Graphoids axioms for Conditional Independence
- D-separation: Inferring CIs in graphs

What we know so far on BN?

- A probability distribution of a Bayesian network having directed graph G, satisfies all the Markov assumptions of independencies.
- 5 graphoid, (or positive) axioms allow inferring more conditional independence relationship for the BN.
- D-separation in G will allow deducing easily many of the inferred independencies.
- G with d-separation yields an I-MAP of the probability distribution.

A Graphical Test of Independence

The inferential power of the graphoid axioms can be tersely captured using a graphical test, known as d-separation, which allows one to mechanically, and efficiently, derive the independencies implied by these axioms.

- To test whether $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$ in DAG G, written $\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, we need to consider every path between a node in $\mathbf{X}$ and a node in $\mathbf{Y}$, and then ensure that the path is blocked by $\mathbf{Z}$.
- The definition of d-separation relies on the notion of blocking a path by a set of variables $\mathbf{Z}$.

$\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ implies $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ for every probability distribution Pr induced by G.

d-speration

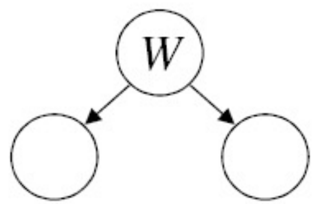
- To test whether **X** and **Y** are d-separated by **Z** in dag G, we need to consider every path between a node in **X** and a node in **Y**, and then ensure that the path is blocked by **Z**.
- A path is blocked by **Z** if **at least** one valve (node) on the path is 'closed' given **Z**.
- A divergent valve or a sequential valve is closed if it is in **Z**
- A convergent valve is closed if it is not on **Z** nor any of its descendants are in **Z**.

d-separation

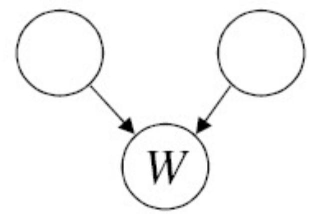
The type of a valve is determined by its relationship to its neighbors on the path.

sequential

divergent

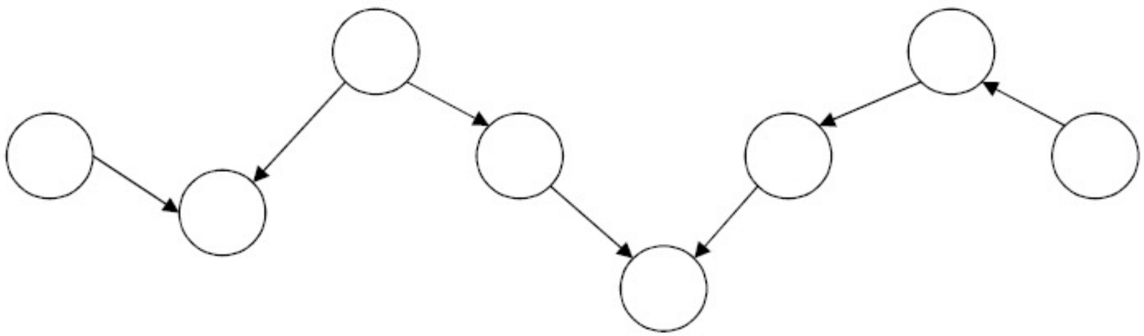


convergent

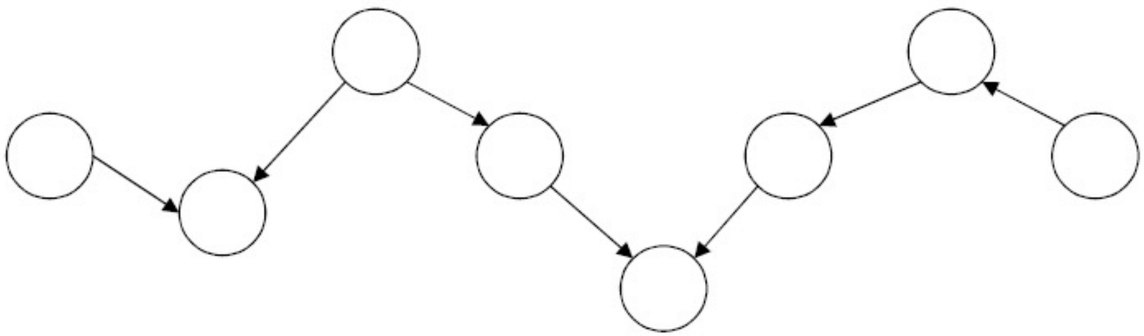


- A sequential valve $\rightarrow W \rightarrow$ arises when W is a parent of one of its neighbors and a child of the other.
- A divergent valve $\leftarrow W \rightarrow$ arises when W is a parent of both neighbors.
- A convergent valve $\rightarrow W \leftarrow$ arises when W is a child of both neighbors.

d-separation



d-separation



Example

A path with 6 valves. From left to right, convergent, divergent, sequential, convergent, sequential, and sequential.

d-separation

Definition

Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ be disjoint sets of nodes in a DAG G. We will say that $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$, written $\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, iff every path between a node in $\mathbf{X}$ and a node in $\mathbf{Y}$ is blocked by $\mathbf{Z}$, where a path is blocked by $\mathbf{Z}$ iff at least one valve on the path is closed given $\mathbf{Z}$.

A path with no valves (i.e., $X \rightarrow Y$) is never blocked.

DEPENDENCE SEMANTICS FOR BAYESIAN NETWORKS

DEFINITION: If X, Y, and Z are three disjoint subsets of nodes in a DAG D, then Z is said to *d-separate* X from Y, denoted $< X \mid Z \mid Y >_D$, if there is no path between a node in X and a node in Y along which the following two conditions hold: (1) every node with converging arrows is in Z or has a descendent in Z and (2) every other node is outside Z.

No path
Is active =
Every path is
blocked

- If a path satisfies the condition above, it is said to be *active*; otherwise, it is said to be *blocked* by Z.

$$<2 \mid 1 \mid 3>_D \quad , \quad \neg <2 \mid 15 \mid 3>_D$$

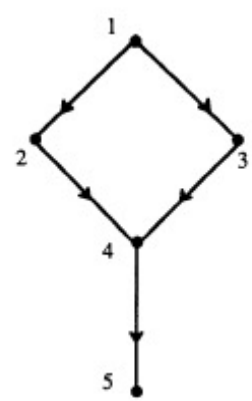
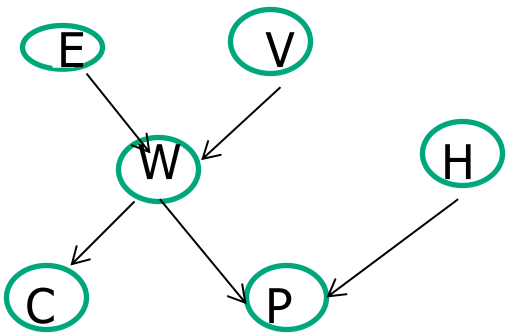


Figure 3.10. A DAG depicting *d-separation*; node 1 blocks the path 2-1-3, while node 5 activates the path 2-4-3.

Bayesian Networks as i-maps

- E: Employment
- V: Investment
- W: Wealth
- H: Health
- C: Charitable contributions
- P: Happiness

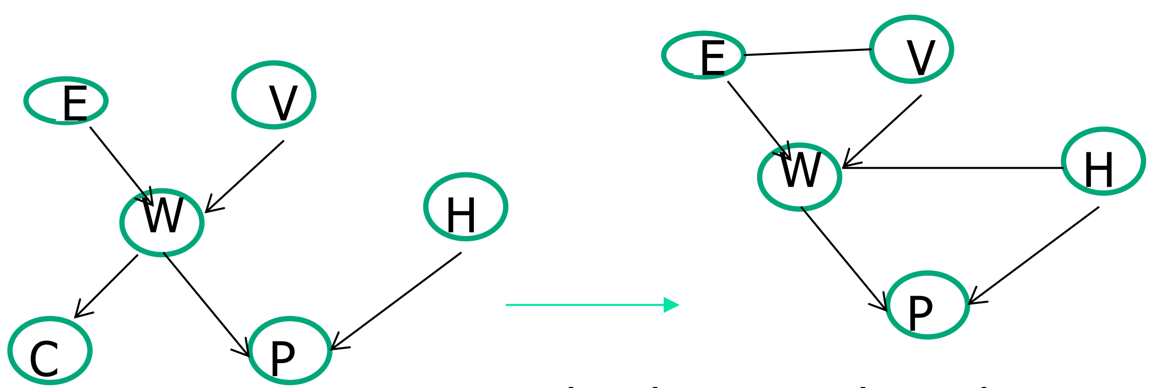


Are C and V d-separated give E and P?
Are C and H d-separated?

d-Seperation Using Ancestral Graph

- X is d-separated from Y given Z (<X,Z,Y>d) iff:
  - Take the ancestral graph that contains **X,Y,Z** and their ancestral subsets.
  - Moralized the obtained subgraph
  - Apply regular undirected graph separation
  - Check: <E,{},V>,<E,P,H>,<C,EW,P>,<C,E,HP>?

<E,P,H>

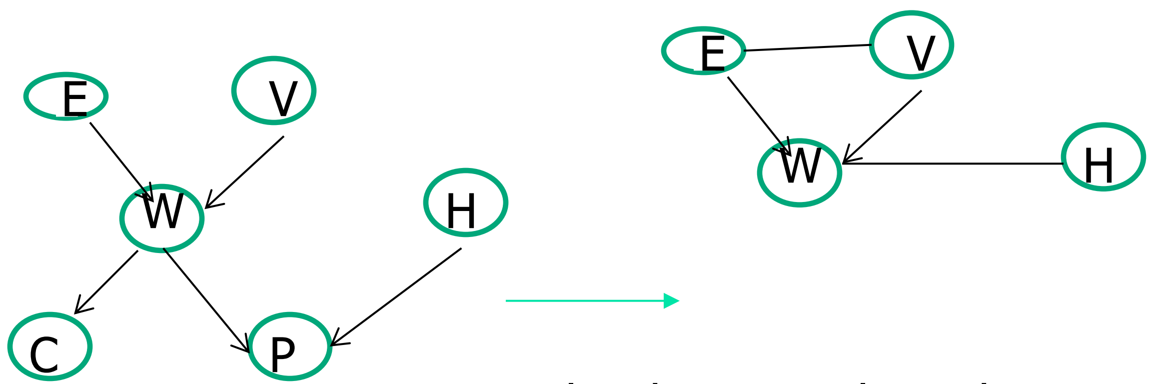


Moralized Ancestral graph

d-Seperation Using Ancestral Graph

- X is d-separated from Y given Z (<X,Z,Y>$_d$) iff:
  - Take the ancestral graph that contains **X,Y,Z** and their ancestral subsets.
  - Moralized the obtained subgraph
  - Apply regular undirected graph separation
  - Check: <E,{},V>,<E,P,H>,<C,EW,P>,<C,E,HP>?

<E,P,H>

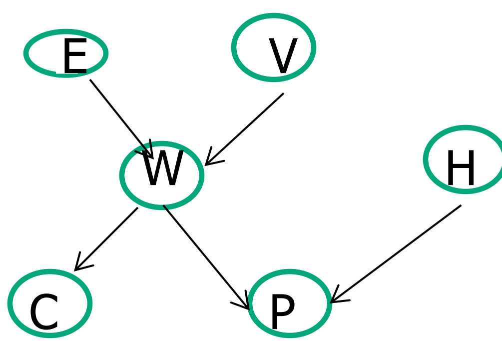


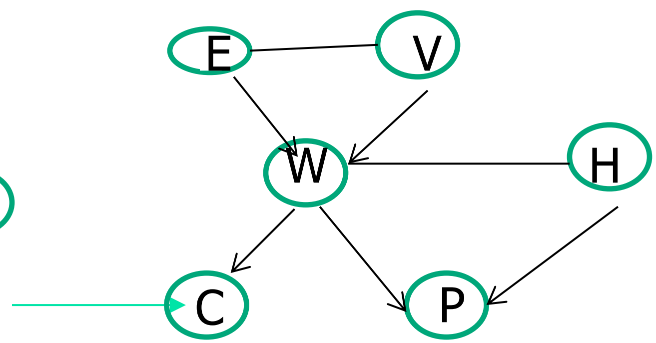
Moralized Ancestral graph

d-Seperation Using Ancestral Graph

- X is d-separated from Y given Z ($<X,Z,Y>_d$) iff:
  - Take the ancestral graph that contains **X,Y,Z** and their ancestral subsets.
  - Moralized the obtained subgraph
  - Apply regular undirected graph separation
  - Check: <E,{},V>,<E,P,H>,<C,EW,P>,<C,E,HP>?

<C,EW,P>



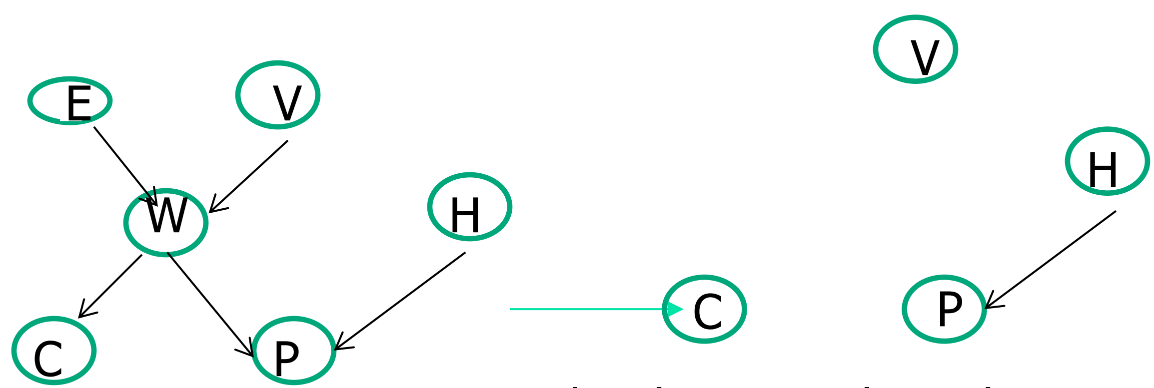


Moralized Ancestral graph

d-Seperation Using Ancestral Graph

- X is d-separated from Y given Z ($\langle X,Z,Y \rangle_d$) iff:
  - Take the ancestral graph that contains **X,Y,Z** and their ancestral subsets.
  - Moralized the obtained subgraph
  - Apply regular undirected graph separation
  - Check: <E,{},V>,<E,P,H>,<C,EW,P>,<C,E,HP>?

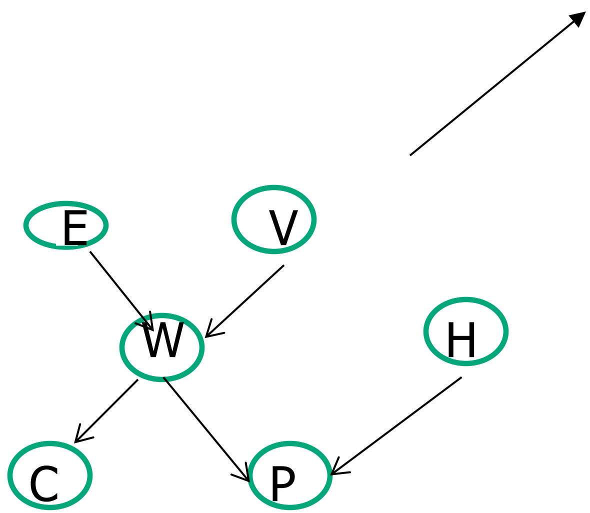
<C,EW,P>



Moralized Ancestral graph

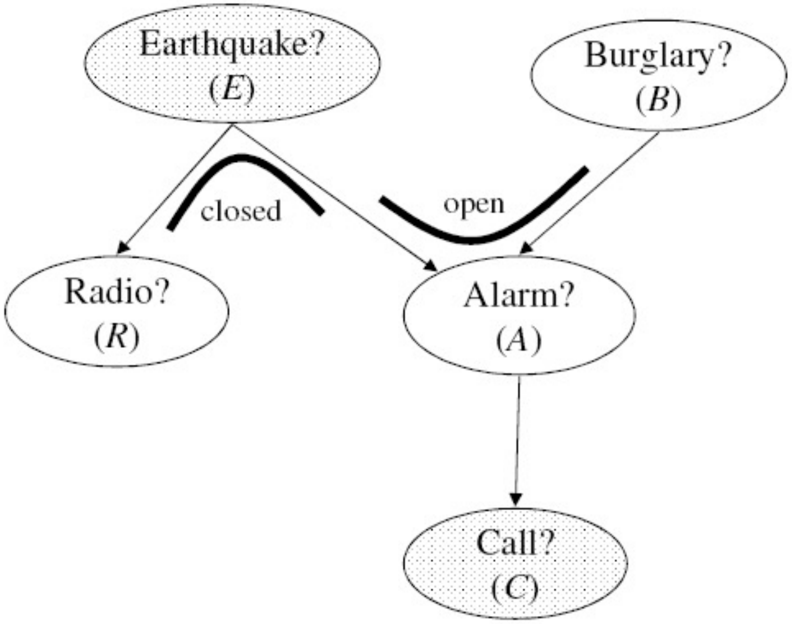
d-Seperation Using Ancestral Graph

- X is d-separated from Y given Z $(<X,Z,Y>_d)$ iff:
  - Take the ancestral graph that contains **X,Y,Z** and their ancestral subsets.
  - Moralized the obtained subgraph
  - Apply regular undirected graph separation
  - Check: <E,{},V>,<E,P,H>,<C,EW,P>,<C,E,HP>?

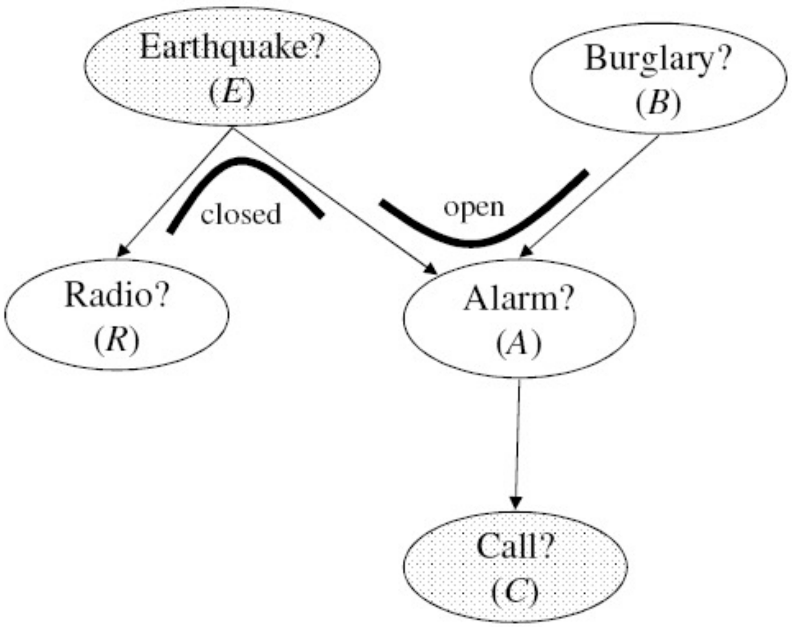


d-separation

$I_{dsep}(R,EC,B)$?



d-separation

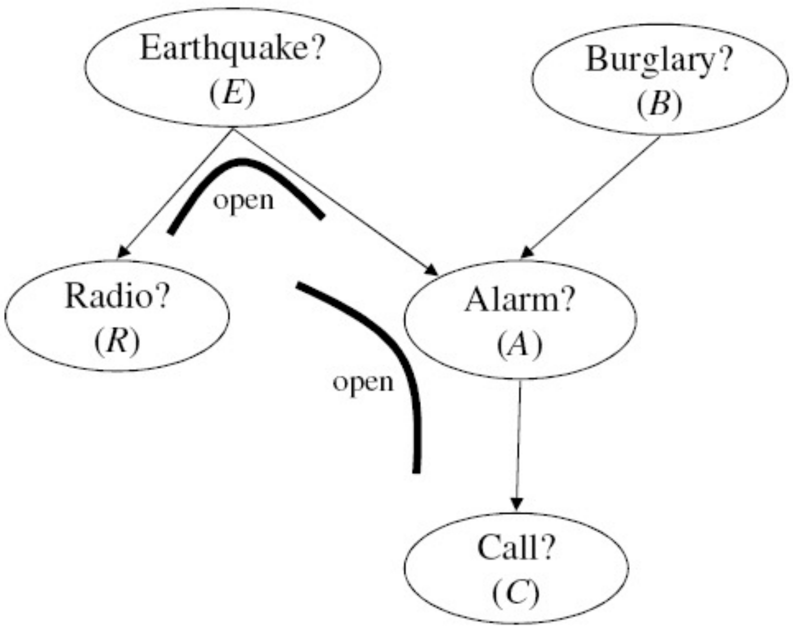


Example

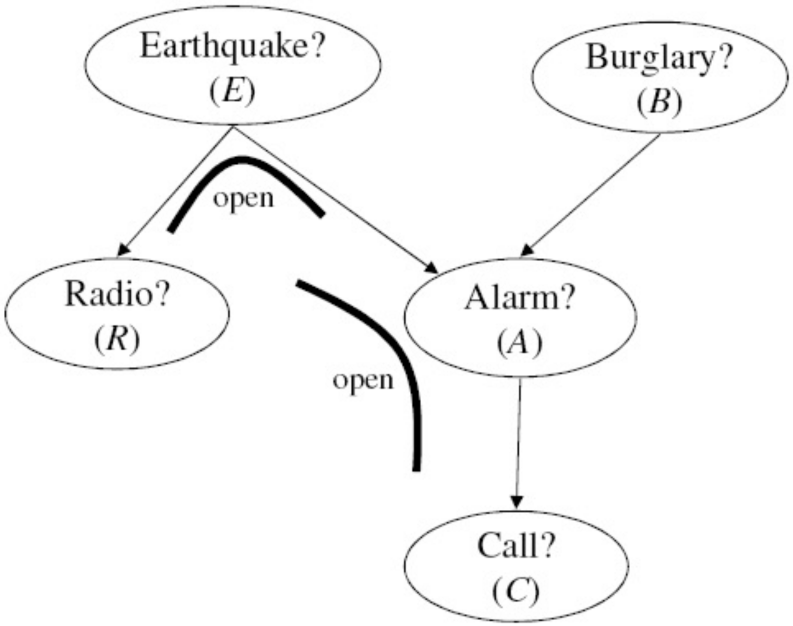
R and B are d-separated by E and C. The closure of only one valve is sufficient to block the path, therefore, establishing d-separation.

d-separation

$I_{dsep}(R,\emptyset,C)$?



d-separation

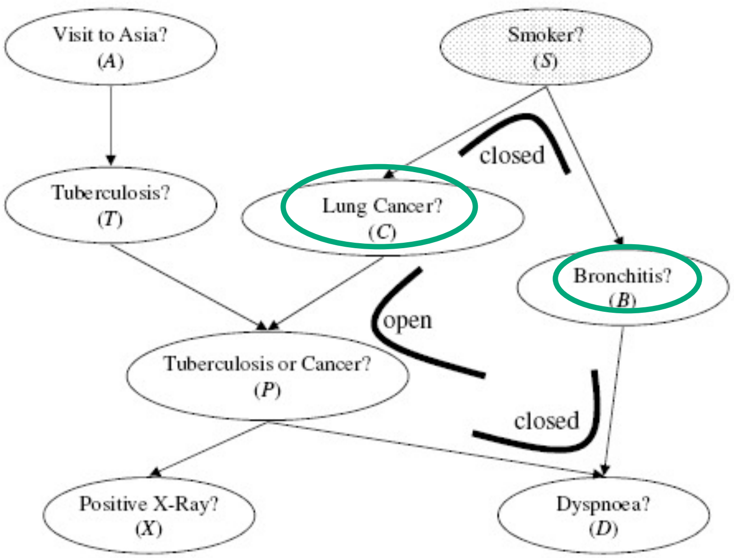


Example

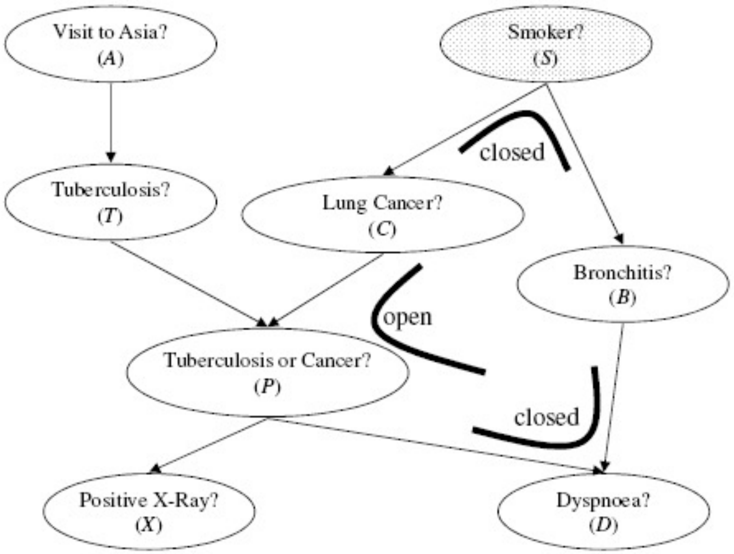
R and C are not d-separated since both valves are open. Hence, the path is not blocked and d-separation does not hold.

d-separation

$I_{dsep}(\mathbf{C},\mathbf{S},\mathbf{B})=?$



d-separation

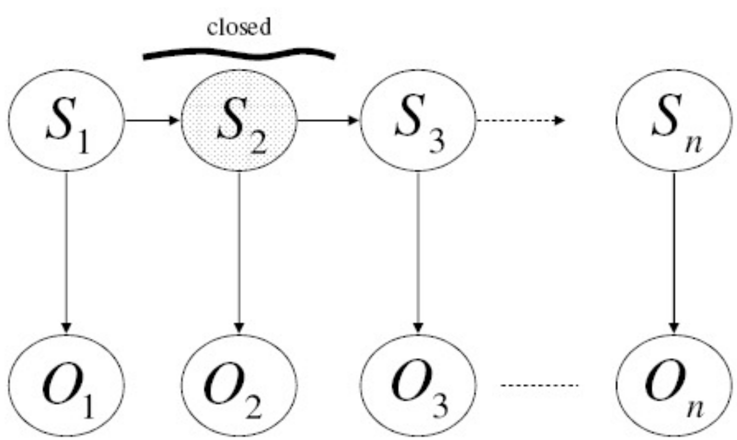


Example

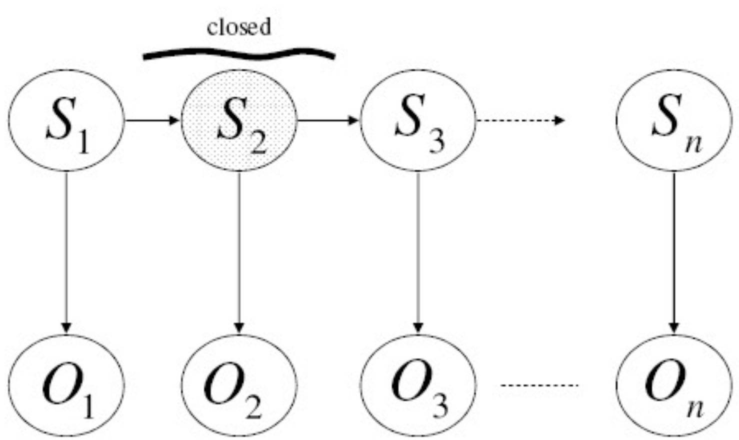
C and B are d-separated by S since both paths between them are blocked by S.

d-separation

Is S1 conditionally on S2 independent of S3 and S4
In the following Bayesian network?



d-separation



Example

Any path between S_1 and $\{S_3, S_4\}$ must have the valve $S_1 \rightarrow S_2 \rightarrow S_3$ on it, which is closed given S_2. Hence, every path from S_1 to $\{S_3, S_4\}$ is blocked by S_2, and we have $\mathrm{dsep}_G(S_1, S_2, \{S_3, S_4\})$, which leads to $I_{\mathrm{Pr}}(S_1, S_2, \{S_3, S_4\})$.

$I_{\mathrm{Pr}}(S_1, S_2, \{S_3, S_4\})$ for any probability distribution Pr which is induced by the DAG.

Outline

- DAGS, Markov(G), Bayesian networks
- Graphoids: axioms of for inferring conditional independence (CI)
- D-separation: Inferring CIs in graphs
  - Soundness, completeness of d-seperation
  - I-maps, D-maps, perfect maps
  - Construction a minimal I-map of a distribution
  - Markov boundary and blanket
  - Markov Networks

Soundness of d-separation

The d-separation test is sound in the following sense.

Theorem

If $\Pr$ is a probability distribution induced by a Bayesian network (G, Θ), then

$$\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y}) \textit{ only if } I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y}).$$

The proof of soundness is constructive, showing that every independence claimed by d-separation can indeed be derived using the graphoid axioms.

Completeness of d-separation

It is not a d-map

d-separation is not complete in the following sense:

- Consider a network with three binary variables $X \rightarrow Y \rightarrow Z$.
- Z is not d-separated from X.
- Z can be independent of X in a probability distribution induced by this network.

Example

Choose the CPT for variable Y so that $\theta_{y|x} = \theta_{y|\bar{x}}$.
Y independent of X since

- $\Pr(y) = \Pr(y|x) = \Pr(y|\bar{x})$ and
- $\Pr(\bar{y}) = \Pr(\bar{y}|x) = \Pr(\bar{y}|\bar{x})$.

Z is also independent of X.

Outline

- DAGS, Markov(G), Bayesian networks
- Graphoids: axioms of for inferring conditional independence (CI)
- D-separation: Inferring CIs in graphs
  - Soundness, completeness of d-seperation
  - I-maps, D-maps, perfect maps
  - Construction a minimal I-map of a distribution
  - Markov boundary and blanket
  - Markov networks

More on DAGs and Independence

Definition

G is an Independence MAP (I-MAP) of $\Pr$ iff every independence declared by d-separation on DAG G holds in the distribution $\Pr$:

$$\text{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y}) \text{ only if } I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y}).$$

Definition

An I-MAP G is minimal if G ceases to be an I-MAP when we delete any edge from G.

By the semantics of Bayesian networks, if $\Pr$ is induced by a Bayesian network (G, Θ), then G must be an I-MAP of $\Pr$, although it may not be minimal.

More on DAGs and Independence

Definition

G is a Dependency MAP (D-MAP) of Pr iff

$$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y}) \text{ only if } \mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y}).$$

If G is a D-MAP of Pr, then the lack of d-separation in G implies a dependence in Pr.

Definition

If DAG G is both an I-MAP and a D-MAP of distribution Pr, then G is called a Perfect MAP (P-MAP) of Pr.

Outline

- DAGS, Markov(G), Bayesian networks
- Graphoids: axioms of for inferring conditional independence (CI)
- D-separation: Inferring CIs in graphs
  - Soundness, completeness of d-seperation
  - I-maps, D-maps, perfect maps
  - Construction a minimal I-map of a distribution
  - Markov boundary and blanket
  - Markov Networks

So how can we construct an I-MAP of a probability distribution?
And a minimal I-Map

Independence MAPs

Given a distribution $\Pr$, how can we construct a DAG G which is guaranteed to be a minimal I-MAP of $\Pr$?

Given an ordering $X_1, \ldots, X_n$ of the variables in $\Pr$:

- Start with an empty DAG G (no edges)
- Consider the variables X_i one by one, for $i = 1, \ldots, n$.
- For each variable X_i, identify a minimal subset $\mathbf{P}$ of the variables in $X_1, \ldots, X_{i-1}$ such that

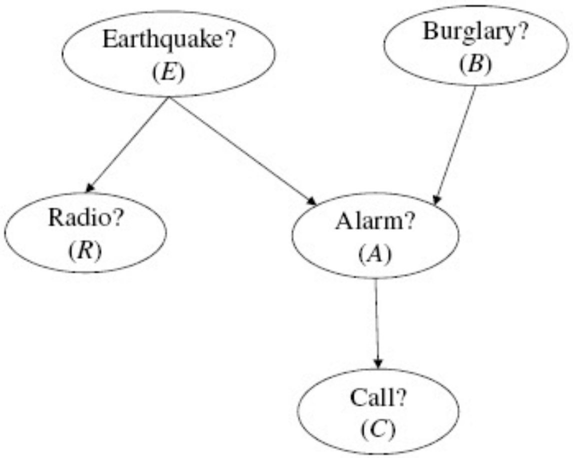
$$I_{\Pr}(X_i, \mathbf{P}, \{X_1, \ldots, X_{i-1}\} \setminus \mathbf{P}).$$

- Make $\mathbf{P}$ the parents of X_i in DAG G.

The resulting DAG is a minimal I-MAP of $\Pr$.

Independence MAPs

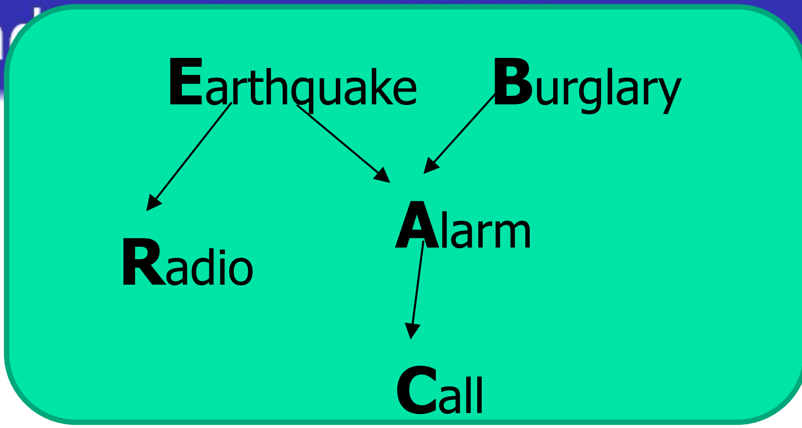
Construct a minimal I-MAP G for some distribution $\Pr$ using the previous procedure and variable order A, B, C, E, R.

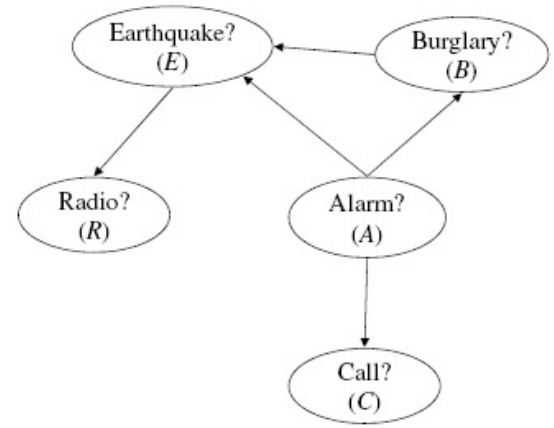


Suppose that DAG G' is a P-MAP of distribution $\Pr$

Independence tests on $\Pr$, $I_{\Pr}(X_i, \mathbf{P}, \{X_1, \ldots, X_{i-1}\} \setminus \mathbf{P})$, can now be reduced to equivalent d-separation tests on DAG G', $\mathrm{dsep}_{G'}(X_i, \mathbf{P}, \{X_1, \ldots, X_{i-1}\} \setminus \mathbf{P})$.

Ind



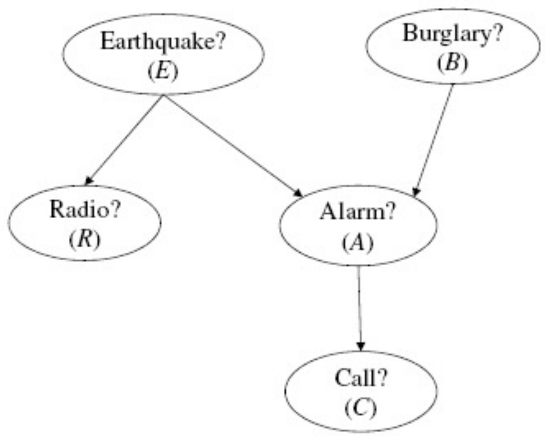


- Variable A added with $\mathbf{P} = \emptyset$.

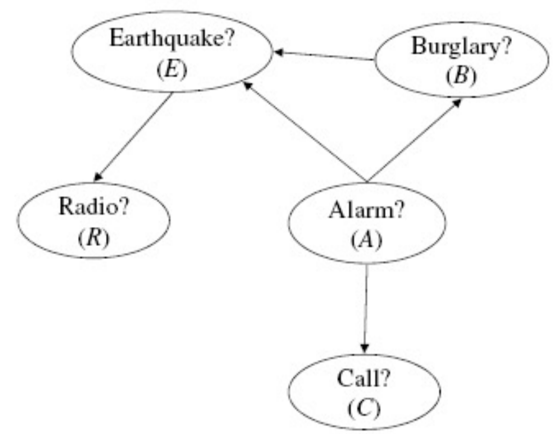
- Variable B added with $\mathbf{P} = A$, since $\mathrm{dsep}_{G'}(B, A, \emptyset)$ holds and $\mathrm{dsep}_{G'}(B, \emptyset, A)$ does not.
- Variable C added with $\mathbf{P} = A$, since $\mathrm{dsep}_{G'}(C, A, B)$ holds and $\mathrm{dsep}(C, \emptyset, \{A, B\})$ does not.
- Variable E added with $\mathbf{P} = A, B$ since this is the smallest subset of A, B, C such that $\mathrm{dsep}_{G'}(E, \mathbf{P}, \{A, B, C\} \setminus \mathbf{P})$ holds.
- Variable R added with $\mathbf{P} = E$ since this is the smallest subset of A, B, C, E such that $\mathrm{dsep}_{G'}(R, \mathbf{P}, \{A, B, C, E\} \setminus \mathbf{P})$ holds.

Independence MAPs

DAG G' and distribution $\Pr$



Minimal I-MAP G



- If $\text{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, then $\text{dsep}_{G'}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$.
- This ceases to hold if we delete any of the five edges in G.

For example, if we delete the edge $E \leftarrow B$, we will have $\text{dsep}_G(E, A, B)$, yet $\text{dsep}_{G'}(E, A, B)$ does not hold.

Independence MAPs

- The minimal I-MAP of a distribution is not unique, as we may get different ones depending on which variable ordering we start with.
- Even when using the same variable ordering, it is possible to arrive at different minimal I-MAPs. This is possible since we may have multiple minimal subsets $\mathbf{P}$ of $\{X_1, \ldots, X_{i-1}\}$ for which $I_{\mathrm{Pr}}(X_i, \mathbf{P}, \{X_1, \ldots, X_{i-1}\} \setminus \mathbf{P})$ holds.
- This can only happen if the probability distribution Pr represents some logical constraints.
- We can ensure the uniqueness of a minimal I-MAP for a given variable ordering if we restrict ourselves to strictly positive distributions.

Perfect Maps for DAGs

- Theorem 10 [Geiger and Pearl 1988]: For any dag D there exists a P such that D is a perfect map of P relative to d-separation.

- Corollary 7: d-separation identifies any implied independency that follows logically from the set of independencies characterized by its dag.

Bayesian Networks as Knowledge-Bases

- Given any distribution, P, and an ordering we can construct a minimal i-map.

- The conditional probabilities of x given its parents is all we need.

- In practice we go in the opposite direction: the parents must be identified by human expert… they can be viewed as direct causes, or direct influences.

BAYESIAN NETWORK AS A KNOWLEDGE BASE

STRUCTURING THE NETWORK

- Given any joint distribution $P(x_1, ..., x_n)$ and an ordering d of the variables in U, Corollary 4 prescribes a simple recursive procedure for constructing a Bayesian network.

- Choose X_1 as a root and assign to it the marginal probability $P(x_1)$ dictated by $P(x_1, ..., x_n)$.

- If X_2 is dependent on X_1, a link from X_1 to X_2 is established and quantified by $P(x_2 | x_1)$. Otherwise, we leave X_1 and X_2 unconnected and assign the prior probability $P(x_2)$ to node X_2.

- At the i-th stage, we form the node X_i, draw a group of directed links to X_i from a parent set Π_{X_i} defined by Eq. (3.27), and quantify this group of links by the conditional probability $P(x_i | \pi_{X_i})$.

- The result is a directed acyclic graph that represents all the independencies that follow from the definitions of the parent sets.

- In practice, $P(x_1,\ldots,x_n)$ is not available.

- The parent sets Π_{X_i} must be identified by human judgment.

- To specify the strengths of influences, assess the conditional probabilities $P(x_i \mid \pi_{X_i})$ by some functions $F_i(x_i, \pi_{X_i})$ and make sure these assessments satisfy

$$\sum_{x_i} F_i(x_i, \pi_{X_i}) = 1 , \tag{3.30}$$

where $0 \le F_i(x_i, \pi_{X_i}) \le 1$

- This specification is complete and consistent because the product form

$$P_a(x_1, \ldots, x_n) = \prod_i F_i(x_i, \pi_{X_i}) \tag{3.31}$$

constitutes a joint probability distribution that supports the assessed quantities.

$$P_a(x_i \mid \pi_{X_i}) = \frac{P_a(x_i, \pi_{X_i})}{P_a(\pi_{X_i})} = \frac{\sum_{x_j \notin (x_i \cup \Pi_{X_i})} P_a(x_1,\ldots,x_n)}{\sum_{x_j \notin \Pi_{X_i}} P_a(x_1,\ldots,x_n)} = F_i(x_i, \pi_{X_i}) \tag{3.32}$$

- DAGs constructed by this method will be called *Bayesian belief networks* or *causal networks* interchangeably.

Outline

- DAGS, Markov(G), Bayesian networks
- Graphoids: axioms of for inferring conditional independence (CI)
- D-separation: Inferring CIs in graphs
  - Soundness, completeness of d-seperation
  - I-maps, D-maps, perfect maps
  - Construction a minimal I-map of a distribution
  - Markov boundary and blanket
  - Markov networks

Blankets and Boundaries

Definition

Let $\Pr$ be a distribution over variables $\mathbf{X}$. A Markov blanket for a variable $X \in \mathbf{X}$ is a set of variables $\mathbf{B} \subseteq \mathbf{X}$ such that $X \notin \mathbf{B}$ and $I_{\Pr}(X, \mathbf{B}, \mathbf{X} \setminus \mathbf{B} \setminus \{X\})$.

A Markov blanket for X is a set of variables which, when known, will render every other variable irrelevant to X.

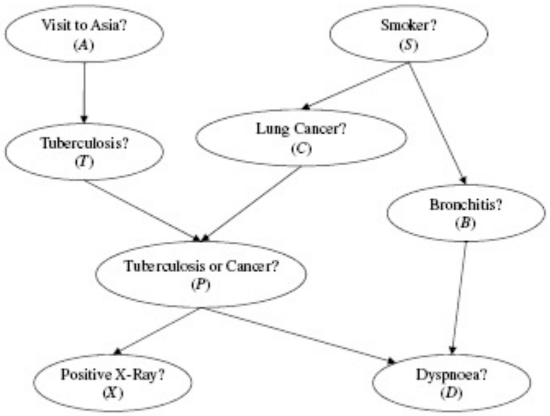
Definition

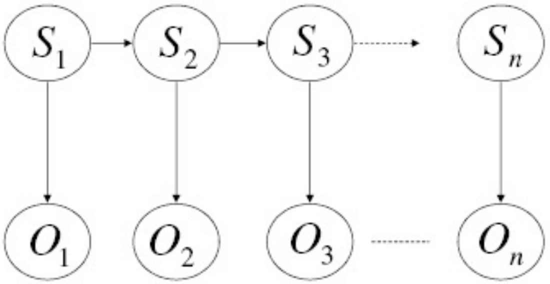
A Markov blanket $\mathbf{B}$ is minimal iff no strict subset of $\mathbf{B}$ is also a Markov blanket. A minimal Markov blanket is a Markov Boundary.

The Markov Boundary for a variable is not unique, unless the distribution is strictly positive.

Blanket Examples

If $\Pr$ is induced by DAG G, then a Markov blanket for variable X with respect to $\Pr$ can be constructed using its parents, children, and spouses in DAG G. Here, variable Y is a spouse of X if the two variables have a common child in DAG G.



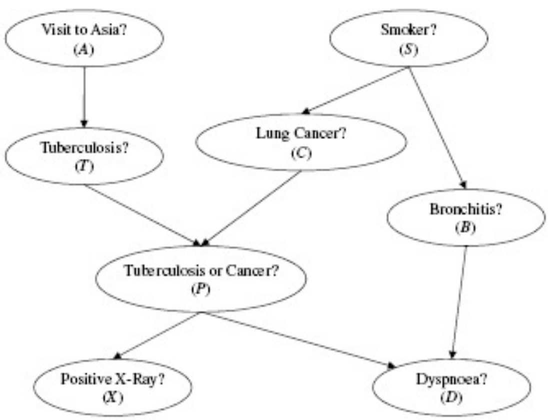


What is a Markov blanket of C?

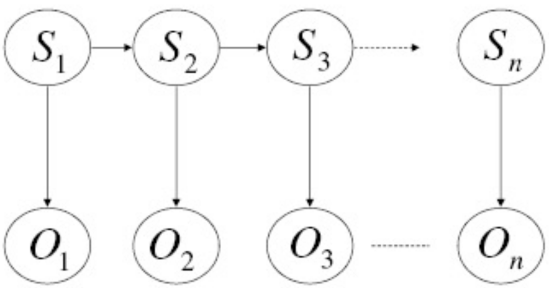
$\{S_{t-1}, S_{t+1}, O_t\}$ is a Markov blanket for every variable S_t, where $t > 1$

Blanket Examples

If Pr is induced by DAG G, then a Markov blanket for variable X with respect to Pr can be constructed using its parents, children, and spouses in DAG G. Here, variable Y is a spouse of X if the two variables have a common child in DAG G.

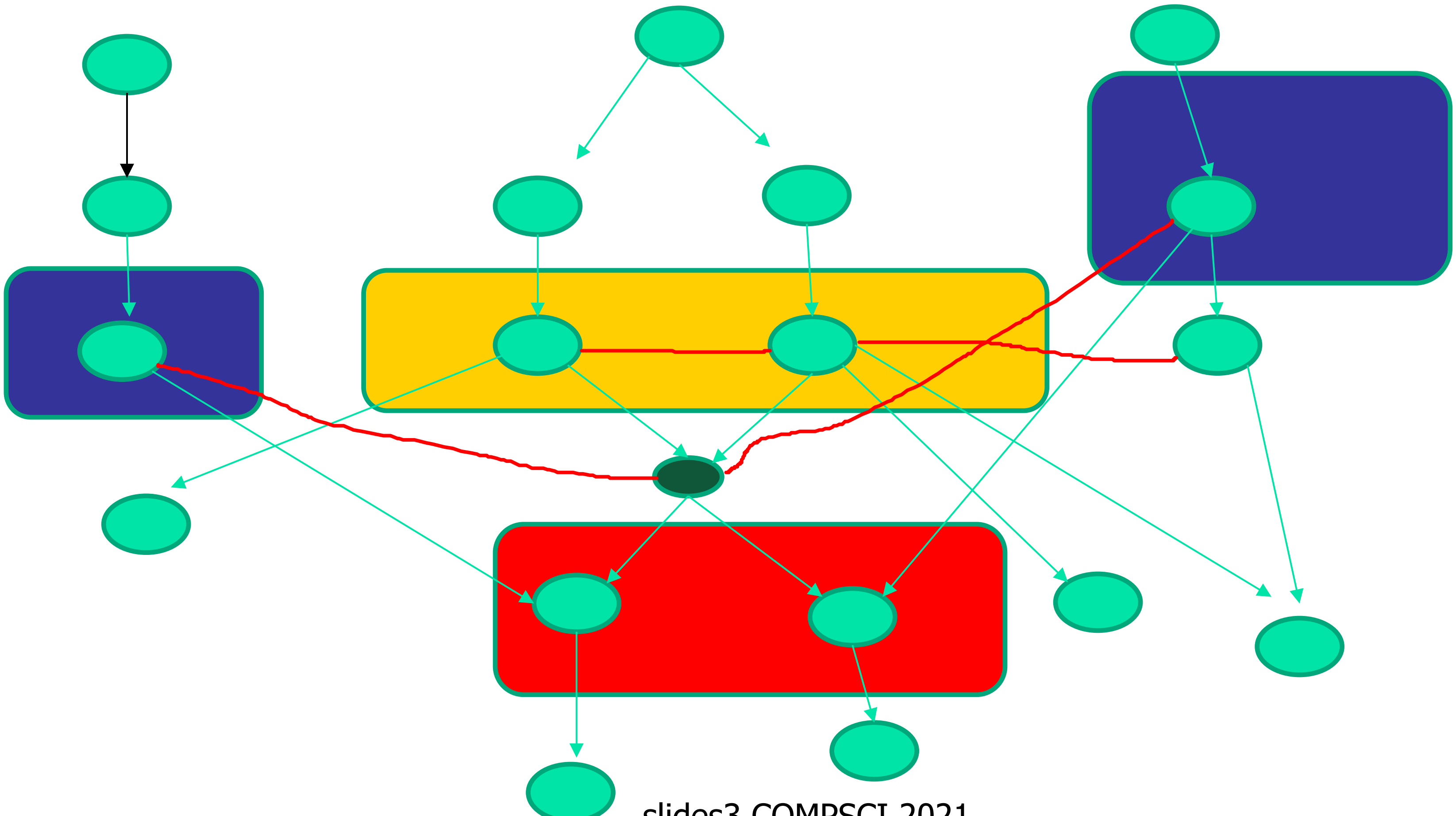


$\{S, P, T\}$ is a Markov blanket for variable C



$\{S_{t-1}, S_{t+1}, O_t\}$ is a Markov blanket for every variable S_t, where $t > 1$

Markov Blanket



Outline

- DAGS, Markov(G), Bayesian networks
- Graphoids: axioms of for inferring conditional independence (CI)
- D-separation: Inferring CIs in graphs
  - Soundness, completeness of d-seperation
  - I-maps, D-maps, perfect maps
  - Construction a minimal I-map of a distribution
  - Markov boundary and blanket
  - Markov networks, Markov Random Fields

Undirected Graphs as I-maps of Distributions

- We say $< X, Z, Y >_G$ iff once you remove Z from the graph X and Y are not connected
- Can we completely capture probabilistic independencies by the notion of separation in a graph?
- Example: 2 coins and a bell.

Graphoids vs Undirected graphs

Graphoids: Conditional Independence

- Symmetry: **I(X,Z,Y) → I(Y,Z,X)**
- Decomposition: **I(X,Z,YW)→ I(X,Z,Y) and I(X,Z,W)**
- Weak union: **I(X,Z,YW)→I(X,ZW,Y)**
- Contraction: **I(X,Z,Y) and I(X,ZY,W)→I(X,Z,YW)**
- Intersection: **I(X,ZY,W) and I(X,ZW,Y) → I(X,Z,YW)**

Seperation in Graphs

Symmetry: **I(X,Z,Y) → I(Y,Z,X)**

Decomposition: **I(X,Z,YW)→ I(X,Z,Y) and I(X,Z,Y)**

Intersection: **I(X,ZW,Y)** and **I(X,ZY,W)→I(X,Z,YW)**

Strong union: I(**X,Z,Y) →** I**(X,ZW, Y)**

Transitivity: I(**X,Z,Y**) → exists t s.t. I(**X,Z**,t) or I(t,**Z,Y**)

See Pearl's book

Markov Networks

- An undirected graph G which is a minimal I-map of a probability distribution Pr, namely deleting any edge destroys its i-mappness relative to (undirected) seperation, is called a **Markov network of P**.

CONCEPTUAL DEPENDENCIES AND THEIR MARKOV NETWORKS

- An agent identifies the following variables as having influence on the main question of being late to a meeting:

  1. The time shown on the watch of Passerby 1.
  2. The time shown on the watch of Passerby 2.
  3. The correct time.
  4. The time it takes to travel to the meeting place.
  5. The arrival time at the meeting place.

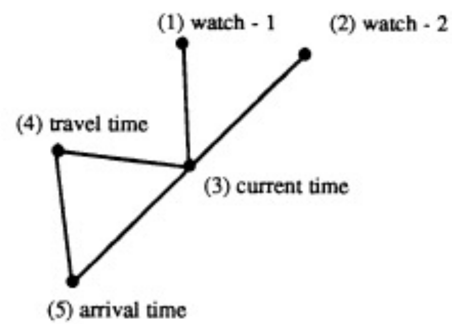
- The construction of G_0 can proceed by one of two methods:
  - The *edge-deletion* method.
  - The *Markov boundary* method.

- The first method requires that for every pair of variables (α, β) we determine whether fixing the values of all other variables in the system will render our belief in α sensitive to β.

- For example, the reading on Passerby 1's watch (1) will vary with the actual time (3) even if all other variables are known, so connect node 1 to node 3

- The Markov boundary method requires that for every variable α in the system, we identify a minimal set of variables sufficient to render the belief in α insensitive to all other variables in the system.

- For instance, once we know the current time (3), no other variable can affect what we expect to read on passerby 1's watch (1).



The unusual edge (3,4) reflects the reasoning that if we fix the arrival time (5) the travel time (4) must depends on current time (3)

Figure 3.6. The Markov network representing the prediction of A's arrival time.

- G_0 can be used as an inference instrument.

  - For example, knowing the current time (3) renders the time on Passerby 1's watch (1) irrelevant for estimating the travel time (4) (i.e., $I(1,3,4)$); 3 is a cutset in G_0, separating 1 from 4.

MARKOV NETWORK AS A KNOWLEDGE BASE

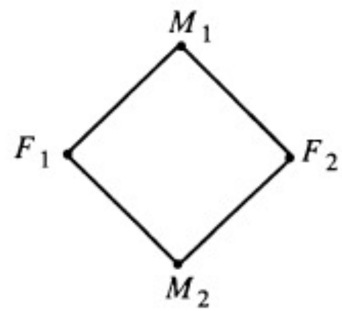


Figure 3.2. An undirected graph representing interactions among four individuals.

How can we construct a probability Distribution that will have all these independencies?

QUANTIFYING THE LINKS

- If couple (M_1, F_2) meet less frequently than the couple (M_1, F_1), then the first link should be weaker than the second
- The model must be consistent, complete and a Markov field of G.
- Arbitrary specification of $P(M_1, F_1)$, $P(F_1, M_2)$, $P(M_2, F_2)$, and $P(F_2, M_1)$ might lead to inconsistencies.
- If we specify the pairwise probabilities of only three pairs, incompleteness will result.

Markov Random Field (MRF)

- A safe method (called *Gibbs' potential*) for constructing a complete and consistent quantitative model while preserving the dependency structure of an arbitrary graph G.

1. Identify the cliques† of G, namely, the largest subgraphs whose nodes are all adjacent to each other.
2. For each clique C_i, assign a nonnegative compatibility function $g_i(\mathbf{c}_i)$, which measures the relative degree of compatibility associated with the value assignment $\mathbf{c}_i$ to the variables included in C_i.
3. Form the product $\prod_i g_i(\mathbf{c}_i)$ of the compatibility functions over all the cliques.
4. Normalize the product over all possible value combinations of the variables in the system

So, How do we learn Markov networks From data?

$$P(x_1, \ldots, x_n) = K \prod_i g_i(\mathbf{c}_i), \tag{3.13}$$

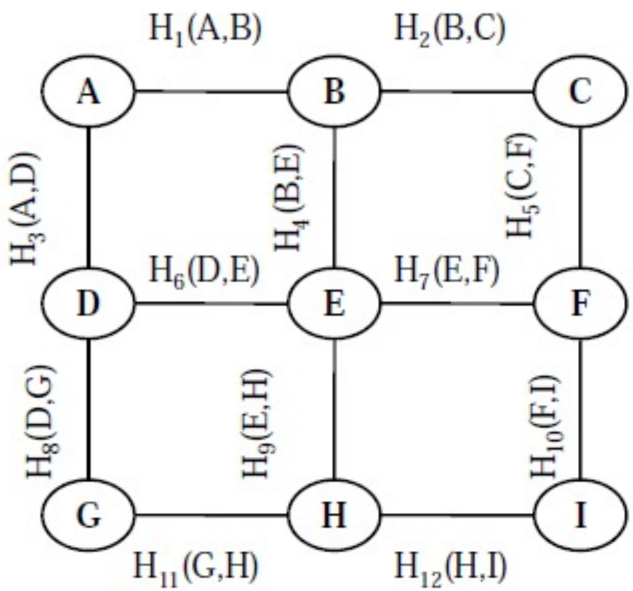
where

$$K = \Big[\sum_{x_1, \ldots, x_n} \prod_i g_i(\mathbf{c}_i) \Big]^{-1}.$$

† We use the term *clique* for the more common term *maximal clique*.

Examples of Bayesian and Markov Networks

Markov Networks



(a)

D	E	H_6(D,E)
0	0	20.2
0	1	12
1	0	23.4
1	1	11.7

(b)

Figure 2.6: (a) An example 3×3 square Grid Markov network (ising model) and (b) An example potential $H_6(D, E)$

network represents a global joint distribution over the variables $\boldsymbol{X}$ *given by:*

$$P(\boldsymbol{x}) = \frac{1}{Z} \prod_{i=1}^{m} H_i(\boldsymbol{x}) \quad , \quad Z = \sum_{\boldsymbol{x} \in \boldsymbol{X}} \prod_{i=1}^{m} H_i(\boldsymbol{x})$$

Sample Applications for Graphical Models

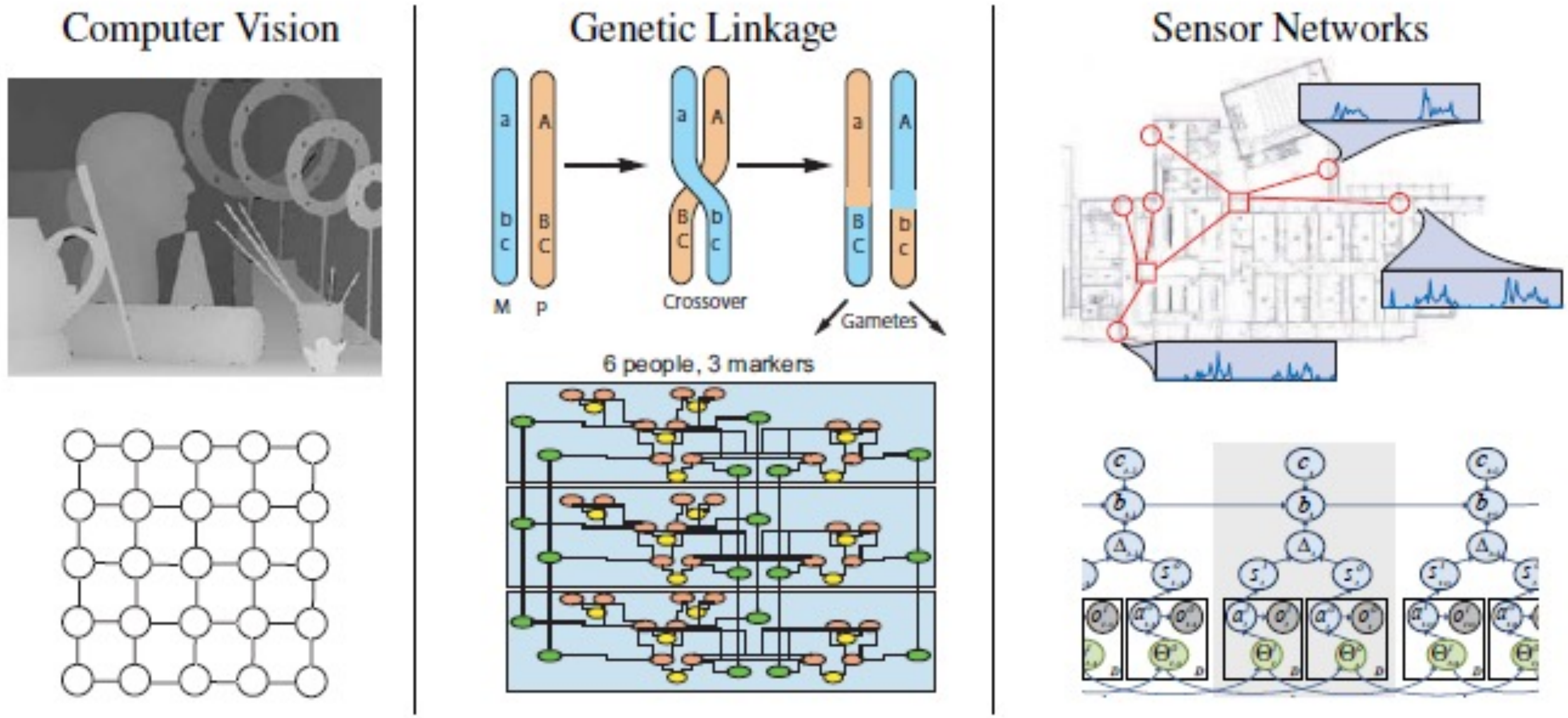


Figure 1: Application areas and graphical models used to represent their respective systems: (a) Finding correspondences between images, including depth estimation from stereo; (b) Genetic linkage analysis and pedigree data; (c) Understanding patterns of behavior in sensor measurements using spatio-temporal models.