Bayesian Networks: Directed Markov Properties

Huizhen Yu

janey.yu@cs.helsinki.fi
Dept. Computer Science, Univ. of Helsinki

Probabilistic Models, Spring, 2010

Outline

Markov Properties on Directed Acyclic Graphs

- Recursive Factorization Property (DF)
- Global Directed Markov Property (DG)
- d-Separation and its Equivalence to (DG)
- Other Directed Markov Properties and Equivalence

Outline

Bayesian Networks and DAG

Bayesian network:

- random variables $X = \{X_v, v \in V\}$,
- a directed acyclic graph $G = (V, E)$, and
- a distribution $P(X)$ that recursively factorizes according to G:

$$p(x) = \prod_{v \in V} p\big(x_v \,|\, x_{\mathrm{pa}(v)}\big), \quad \text{where } \ \mathrm{pa}(v) = \{\text{parents of } v \text{ in } G\}. \tag{1}$$

This property of P will be denoted (DF). ('D' stands for 'directed.')

(DF) is a well-defined property:

- The right-hand side of (1) defines a valid distribution, i.e., the sum over all $x_v, v \in V$ equals 1. (This is obvious, because G has no cycles.)
- There exists P that satisfies the property.

  This also seems obvious. To argue more formally, we introduce well-orderings of vertices.

Well-Ordering and Property (DF)

We can number the vertices of G in such a way that

$$(\alpha, \beta) \in E \quad \Rightarrow \quad \text{number}(\alpha) < \text{number}(\beta).$$

(I.e., a child has a larger number than any of its parents.)

Any such ordering is called a *well-ordering*.

We then define the set $\text{pr}(\beta)$ of *predecessors* of β to be

$$\text{pr}(\beta) = \{\alpha \mid \text{number}(\alpha) < \text{number}(\beta)\}.$$

Using this ordering, we can express any distribution $P(X)$ as

$$p(x) = p(x_1)p(x_2 \,|\, x_1) \cdots p(x_n \,|\, x_1, \ldots, x_{n-1}), \quad n = |V|. \tag{2}$$

So, if P satisfies the conditional independence relations

$$p(x_i \,|\, x_1, \ldots, x_{i-1}) = p(x_i \,|\, x_{\text{pa}(i)}), \quad \text{i.e.,} \quad X_v \perp X_{\text{pr}(v)} \mid X_{\text{pa}(v)}, \tag{3}$$

then Eq. (2) reduces to (1), implying that P satisfies (DF).

The property in (3) is called the *ordered directed Markov property* (DO).

This shows there exists P that satisfies (DF), and (DO) $\Rightarrow$ (DF).

(DF), (F) and Moral Graph

If P factorizes recursively according to G, then P factorizes according to an undirected graph G^m:

$$p(x) = \prod_{v \in V} p\big(x_v \,|\, x_{\mathrm{pa}(v)}\big) \quad \Rightarrow \quad p(x) = \prod_{C \in \mathcal{C}(G^m)} \phi_C(x_C)$$

for some functions ϕ_C, where

- $\mathcal{C}(G^m)$ denotes the set of complete subsets of G^m; and
- G^m is constructed by making v and $\mathrm{pa}(v)$ a complete subset for every $v \in V$.

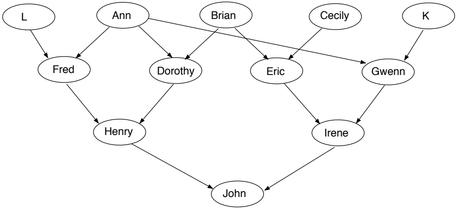
The construction of G^m is identical to modifying G by

- "marrying" the parents – adding undirected edges between all pairs of parents who have a common child but are not linked by an edge in G, and then
- dropping the arrows of all directed edges.

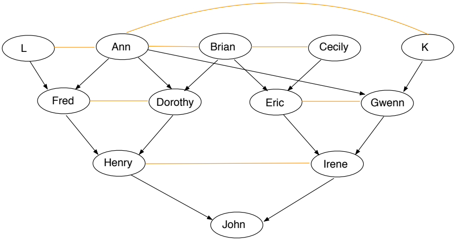
G^m is called the *moral graph* of G.

Moral Graph for Example Stud-Farm

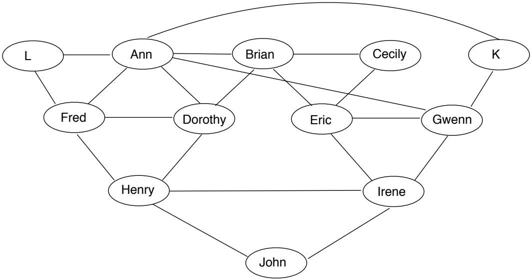
Geneological structure for the horses:



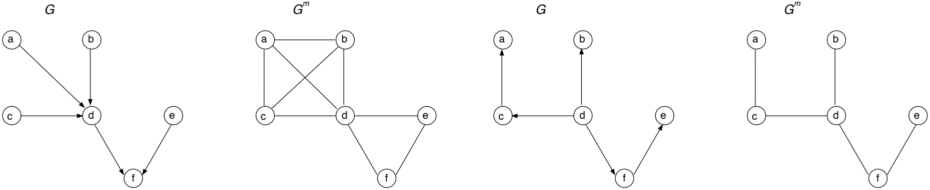
"Marrying" the parents:



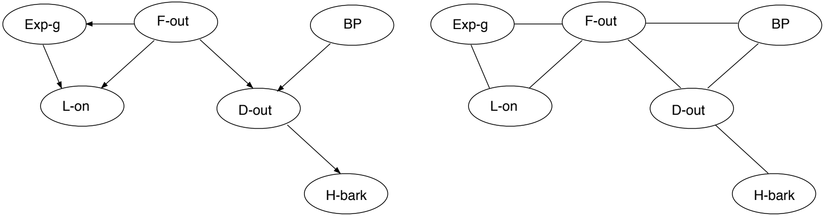
G^m:



Examples of Moral Graphs



Example Family-Out?:



Independence Relations Implied by G^m

Since P factorizes according to G^m, P satisfies the global (G), local (L), and pairwise (P) Markov properties with respect to G^m. (See Lecture 5.)

So we can read off some independence relations from G^m using (G).

For example, consider (L): Denote $\text{ne}(v) = \{\text{neighbors of } v \text{ in } G^m\}$. Then,

$$X_v \perp X_{V \setminus \text{ne}(v)} \,|\, X_{\text{ne}(v)}. \tag{4}$$

Who are these neighbors $\text{ne}(v)$, viewed in G?

- Edges of G^m are either from G or from "marriages," which means

  $$\text{ne}(v) = \text{pa}(v) \cup \text{ch}(v) \cup \{w \,|\, \text{ch}(w) \cap \text{ch}(v) \neq \emptyset\},$$

  where $\text{ch}(v) = \{\text{children of } v \text{ in } G\}$.
- $\text{ne}(v)$ is called the *Markov blanket* of v in G, denoted by $\text{bl}(v)$.

So the statement that given its Markov blanket, X_v is independent of the rest of the variables,

$$X_v \perp X_{V \setminus \text{bl}(v)} \,|\, X_{\text{bl}(v)},$$

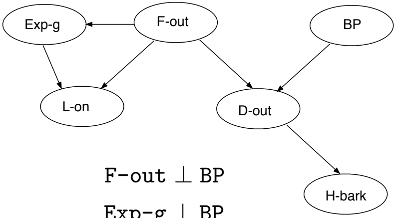
is just a rewrite of the local Markov property (4) with respect to G^m.

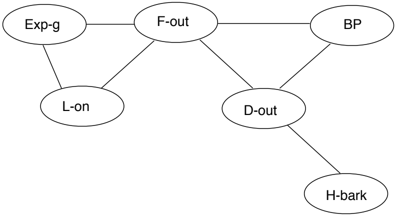
Independence Relations Implied by G^m

Many independence relations in P are not captured by G^m, however. This is because P has extra properties not possessed by a general distribution that factorizes according to the undirected graph G^m:

$$p(x) = \prod_{v \in V} p\left(x_v \,|\, x_{\mathrm{pa}(v)}\right) \quad \not\Leftarrow \quad p(x) = \prod_{C \in \mathcal{C}(G^m)} \phi_C(x_C).$$

Example Family-out?:





$\texttt{F-out} \perp \texttt{BP}$

$\texttt{Exp-g} \perp \texttt{BP}$

Clearly, without observing anything, whether the family is out, or whether guests are being expected, is independent of the dog having bowel problem. But these vertices are connected in G^m. For their dependence relations, we need to look for structures in their marginal distributions.

Outline

Ancestral Sets

A sequence of distinct vertices $\alpha_1, \ldots, \alpha_m$ is called a *path* in G, if $(\alpha_{i-1}, \alpha_i) \in E$, $i = 2, \ldots, m$.

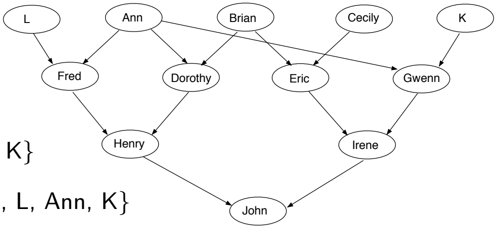
For two vertices α, β of G, we say α is an *ancestor* of β and β a *descendant* of α, if there is a path from α to β in G. Denote by $\mathrm{an}(\alpha)$ the set of *ancestors* of α and by $\mathrm{de}(\alpha)$ the set of *descendants* of α.

We say A is an *ancestral set* if $\mathrm{pa}(v) \subseteq A, \forall v \in A$.

Let $\mathrm{An}(A)$ denote the *minimal* ancestral set containing A.

(In a DAG, $\mathrm{An}(A) = A \cup_{\alpha \in A} \mathrm{an}(\alpha)$.)

Example Stud-Farm:



$\mathrm{de}(\text{Irene}) = \{\text{John}\}$

$\mathrm{an}(\text{Irene})$
$= \{\text{Eric, Gwenn, Brian, Cecily, Ann, K}\}$

$\mathrm{An}(\{\text{Fred, Gwenn}\}) = \{\text{Fred, Gwenn, L, Ann, K}\}$

Factorization of Marginals corresponding to Ancestral Sets

For $A \subseteq V$, let G_A denote the subgraph of G on A, i.e.,

$$G_A = (A, E_A), \quad E_A = E \cap (A \times A),$$

and let P_A denote the marginal distribution of X_A.

Fact: *For an ancestral set A, P_A factorizes recursively according to G_A.*

Verifying the fact: We have

$$p(x) = \prod_{v \in V} p\big(x_v \,|\, x_{\mathrm{pa}(v)}\big) = \Big(\prod_{v \in A} p\big(x_v \,|\, x_{\mathrm{pa}(v)}\big)\Big) \cdot \Big(\prod_{v \in V \setminus A} p\big(x_v \,|\, x_{\mathrm{pa}(v)}\big)\Big).$$

Since A is ancestral, $\mathrm{pa}(v) \subseteq A, \forall v \in A$, so the first term is a function of x_A only. Hence

$$\begin{aligned} p_A(x_A) &= \sum_{x_{V\setminus A}} p(x) = \Big(\prod_{v \in A} p\big(x_v \,|\, x_{\mathrm{pa}(v)}\big)\Big) \cdot \sum_{x_{V\setminus A}} \Big(\prod_{v \in V \setminus A} p\big(x_v \,|\, x_{\mathrm{pa}(v)}\big)\Big) \\ &= \Big(\prod_{v \in A} p\big(x_v \,|\, x_{\mathrm{pa}(v)}\big)\Big) \cdot 1, \end{aligned}$$

where $\sum_{x_{V\setminus A}} \prod_{v \in V\setminus A} p\big(x_v \,|\, x_{\mathrm{pa}(v)}\big) = 1$ can be seen by taking any well-ordering of vertices and summing over $x_v, v \notin A$ one by one in the descending order, in other words, successively summing over variables that have no children left.

Global Directed Markov Property

Let A, B, S be three disjoint subsets of V.

Based on the fact in the preceding slide and the implications of the recursive factorization property shown in slides 6 and 9,

$$P_{\mathrm{An}(A\cup B\cup S)} \text{ factorizes recursively according to } G_{\mathrm{An}(A\cup B\cup S)},$$
$$\Downarrow$$
$$P_{\mathrm{An}(A\cup B\cup S)} \text{ factorizes according to } \left(G_{\mathrm{An}(A\cup B\cup S)}\right)^m,$$
$$\Downarrow$$
$$P_{\mathrm{An}(A\cup B\cup S)} \text{ obeys the global Markov property on } \left(G_{\mathrm{An}(A\cup B\cup S)}\right)^m.$$

Therefore,

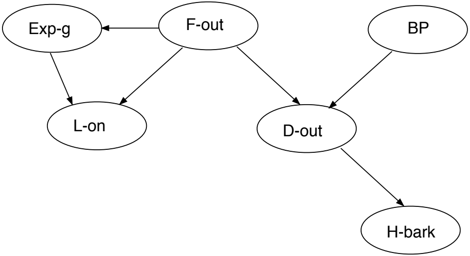
$$A \perp B \,|\, S \quad \text{in} \quad \left(G_{\mathrm{An}(A\cup B\cup S)}\right)^m \quad \Rightarrow \quad X_A \perp X_B \,|\, X_S. \tag{5}$$

The property in (5) is called the *global directed Markov property* (DG).

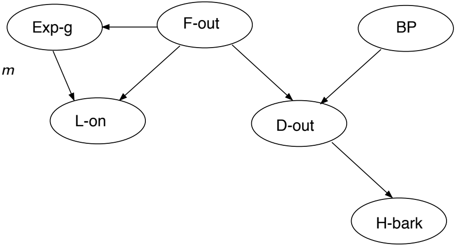
Illustrations of (DG)

Example Family-out?:

(1) Is $\texttt{F-out} \perp \texttt{BP}$? (2) Is $\texttt{Exp-g} \perp \texttt{BP}$? (3) Is $\texttt{L-on} \perp \texttt{BP} \mid \texttt{D-out}$?

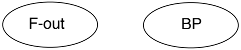


Answers to Previous Questions



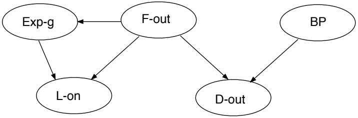
(1) Is $\texttt{F-out} \perp \texttt{BP}$? – Yes.

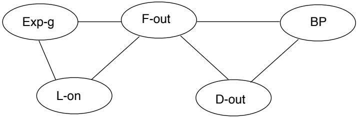
Both $G_{\mathrm{An}(A\cup B)}$ and $\left(G_{\mathrm{An}(A\cup B)}\right)^m$ are



(3) Is $\texttt{L-on} \perp \texttt{BP} \mid \texttt{D-out}$? – No.

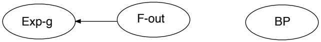
$G_{\mathrm{An}(A\cup B\cup S)}$ and $\left(G_{\mathrm{An}(A\cup B\cup S)}\right)^m$ are

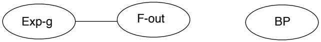




(2) Is $\texttt{Exp-g} \perp \texttt{BP}$? – Yes.

$G_{\mathrm{An}(A\cup B)}$ and $\left(G_{\mathrm{An}(A\cup B)}\right)^m$ are





Outline

Markov Properties on Directed Acyclic Graphs

Recursive Factorization Property (DF)

Global Directed Markov Property (DG)

d-Separation and its Equivalence to (DG)

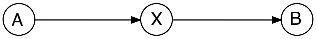
Other Directed Markov Properties and Equivalence

d-Separation: Intuition

Judea Pearl introduced the concept *d-separation* to define the Markov property on directed graphs. The intuition comes from how evidence propagates in the network. To explain this, we consider three types of edge connections at a vertex X, as illustrated below.

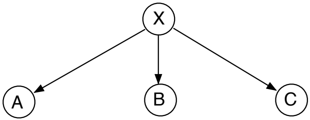
What if X is *instantiated*?

- Serial connection:

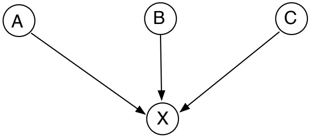


$$A \perp B \,|\, X.$$

- Diverging connection:
  A, B, C are mutually independent given X.



- Converging connection:
  A, B, C are mutually independent if X is not observed, but become dependent given X.

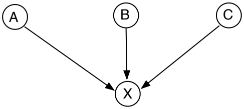


So when X is instantiated, in the first two cases, it blocks the transmission of evidence from one linked vertex to another, while in the third case, it opens the channel for transmission of evidence.

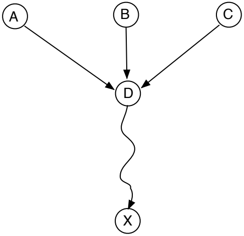
But if X is unobserved, the situation is the opposite. (remark added after lecture)

d-Separation: Intuition

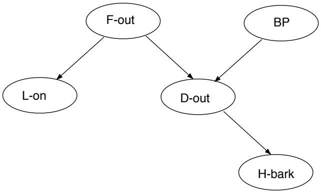
Similarly, in the case of a converging connection, when X is not instantiated ("hard" evidence) but receives "soft" evidence about its value, it also opens the channel for the evidence of any linked vertex to pass.



So is the case when there is a path to X from the vertex at the converging point.



Example Family-Out?:



Hearing dog barking makes any change in the belief of family-out to affect the belief of the dog having bowel trouble. For example, if the family is out, then "bowel trouble" would be explained away.

d-Separation: Definition

d-Separation is defined in terms of the non-existence of certain kind of routes, called *active trails*, which can open channels for evidence transmission in the graph.

Definitions:

A sequence of distinct vertices $\alpha_1, \alpha_2, \ldots, \alpha_m$ is called a *trail* in G, if for $i = 2, \ldots, m$, either $(\alpha_{i-1}, \alpha_i) \in E$ or $(\alpha_i, \alpha_{i-1}) \in E$. I.e., a trail is a path in the undirected version of G.

Given $S \subseteq V$, a trail is said to be *active*, if it satisfies two conditions:

(1) Every vertex with converging arrows is in S or has a descendant in S;

(2) Every other vertex is outside S. Such vertex is on a serial or diverging connection; if instantiated (in S), it would block the trail.

Trails that are not active are said to be blocked by S. (remark added after lecture)

d-Separation: Let A, B, S be three disjoint subsets of vertices of a DAG G. S is said to *d-separate* A from B, if there are *no* active trails from A to B.

A Markov property defined in terms of d-separation:

$$S \text{ d-separates } A \text{ from } B \quad \Rightarrow \quad X_A \perp X_B \mid X_S.$$

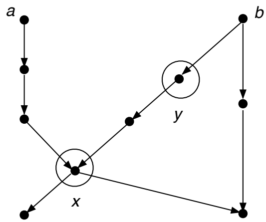
d-Separation: Illustration

Recall the definition of an active trail:

(1) Every vertex with converging arrows is in S or has a descendant in S;

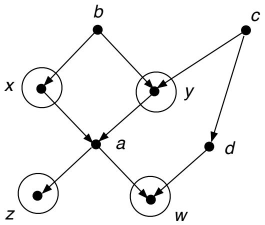
(2) Every other vertex is outside S.

Does $\{x, y\}$ d-separate a from b?



Two trails from a to b here,
one blocked by x and the other by y.

Does $\{x, y, w, z\}$ d-separate a from b or c or d?



The d-separation concept is not convenient for computation, as can be seen from the above examples.

Equivalence between (DG) and d-Separation Based Markov Property

Proposition: *Let A, B, S be three disjoint subsets of vertices of a DAG G. Then S d-separates A from B in G if and only S separates A from B in* $\left(G_{An(A \cup B \cup S)}\right)^m$.

Proof sketch: By definition,

S does not d-separate A from B in G. $\Leftrightarrow$ $\exists$ an active trail from A to B in G.

We will show: (a) $\Downarrow$ and (b) $\Uparrow$

S does not separate A from B in $\left(G_{\mathrm{An}(A \cup B \cup S)}\right)^m$. $\Leftrightarrow$ $\exists$ a path from A to B in $\left(G_{\mathrm{An}(A \cup B \cup S)}\right)^m$, circumventing S.

Recall the definition of an active trail:

(1) Every vertex with converging arrows is in S or has a descendant in S;

(2) Every other vertex is outside S.

Equivalence Proof: Part (a)

Under the assumption, there is an active trail from A to B in G. We want to show that there is a path from A to B in $\left(G_{\mathrm{An}(A \cup B \cup S)}\right)^m$, circumventing S.

We first argue that

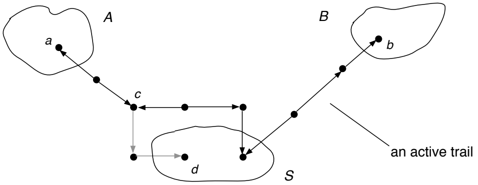
- All the vertices on the trail are in $\mathrm{An}(A \cup B \cup S)$:

  Since the trail is active, vertices with converging arrows are in $\mathrm{An}(S)$.

  Other vertices are on either a serial or diverging connection. Consider any one of these vertices and call it γ. Following an outgoing edge from γ, either the trail leads all the way to A or B without traversing against the direction of any edge on its way, or it leads to a vertex with converging arrows. In the former case, $\gamma \in \mathrm{An}(A)$ or $\mathrm{An}(B)$, and in the latter case, $\gamma \in \mathrm{An}(S)$.

This shows that the trail lies in $\left(G_{\mathrm{An}(A \cup B \cup S)}\right)^m$.

Illustration:

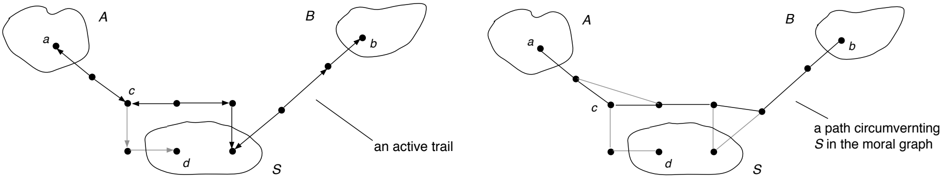


Equivalence Proof: Part (a) Cont'd

We now modify the trail, if necessary, to make it circumvent S in $\left(G_{\mathrm{An}(A\cup B\cup S)}\right)^m$.

Consider the vertices on the trail. Since the trail is active, any vertex that is in S must have converging arrows. The marriage edge it creates between its two parents on the trail are in $\left(G_{\mathrm{An}(A\cup B\cup S)}\right)^m$, and the parents are not in S (since they cannot have converging arrows). This shows that we can create the desired path from A to B in $\left(G_{\mathrm{An}(A\cup B\cup S)}\right)^m$, by following the trail and traversing along the marriage edges to circumvent S whenever necessary.

Illustration:



Equivalence Proof: Part (b)

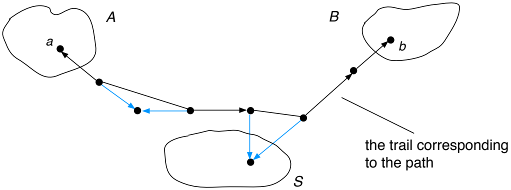
Under the assumption, there is a path from A to B in $\left(G_{\mathrm{An}(A\cup B\cup S)}\right)^m$, circumventing S. We want to construct an active trail from A to B in G.

First, from the given path in $\left(G_{\mathrm{An}(A\cup B\cup S)}\right)^m$,

- We construct a trail from A to B in $G_{\mathrm{An}(A\cup B\cup S)}$:

  If the path traverses through a marriage edge, the child of the marriage must be in $\mathrm{An}(A \cup B \cup S)$. We modify the path to go through that child, thereby reducing the number of marriage edges of the path by one. Repeating this process, we obtain a path in $\left(G_{\mathrm{An}(A\cup B\cup S)}\right)^m$ which does not have marriage edges, and which corresponds to a trail from A to B in $G_{\mathrm{An}(A\cup B\cup S)}$.

Illustration of removing marriage edges from the path:



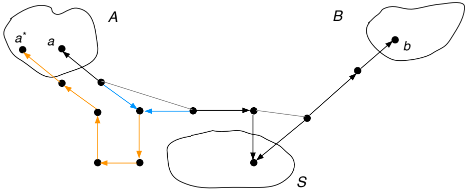
Equivalence Proof: Part (b) Cont'd

We now modify the trail just constructed to make it active. Consider the vertices of the trail. By construction, all the vertices that are not in the original path have converging arrows, and since the original path circumvents S, all the vertices with non-converging arrows are outside S. So, to satisfy the first condition in the definition of an active trail, we only need to consider those vertices with converging arrows. Consider any one of them and call it γ.

- Case (i): $\gamma \in \mathrm{An}(S)$. Then, it does not block the trail.
- Case (ii): $\gamma \notin \mathrm{An}(S)$. In this case, $\gamma \in \mathrm{An}(A \cup B)$ (since $\gamma \in \mathrm{An}(A \cup B \cup S)$). Therefore, there is a path in $G_{\mathrm{An}(A \cup B \cup S)}$ from γ to its descendant in either the set A or B, and we use this path to replace the piece of the trail from γ to the corresponding set. This eliminates a converging connection from the trail.

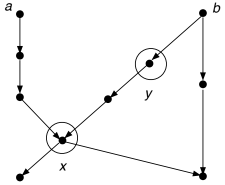
Repeating the above process, eventually all the vertices on the trail that have converging arrows must be in $\mathrm{An}(S)$, and the trail is made active.

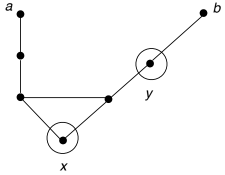
Illustration of making an active trail:



Illustration

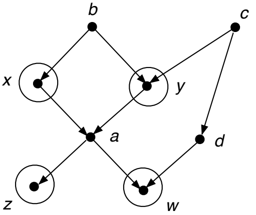
Does $\{x, y\}$ d-separate a from b? – Yes.

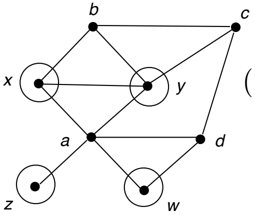




$$\left(G_{\mathrm{An}\left(\{a,b,x,y\}\right)}\right)^m$$

Does $\{x, y, w, z\}$ d-separate a from b or c or d? – No.





$$\left(G_{\mathrm{An}\left(\{a,x,y,w,z\}\right)}\right)^m$$

Outline

Markov Properties on Directed Acyclic Graphs

- Recursive Factorization Property (DF)
- Global Directed Markov Property (DG)
- d-Separation and its Equivalence to (DG)
- Other Directed Markov Properties and Equivalence

Local Directed Markov Property

Denote $\mathrm{nd}(v)$ the set of *non-descendants* of v in G, i.e., $\mathrm{nd}(v) = V \setminus (\mathrm{de}(v) \cup \{v\})$.

Consider v, its parents $\mathrm{pa}(v)$, and its non-descendants $\mathrm{nd}(v)$.

- Because $\mathrm{An}\big(\{v\} \cup \mathrm{pa}(v) \cup \mathrm{nd}(v)\big)$ is the minimal ancestral set containing these vertices,

$$\mathrm{ch}(v) \cap \mathrm{An}\big(\{v\} \cup \mathrm{pa}(v) \cup \mathrm{nd}(v)\big) = \emptyset.$$

- Edges in the moral graph $\big(G_{\mathrm{An}(\{v\} \cup \mathrm{pa}(v) \cup \mathrm{nd}(v))}\big)^m$ are either from $G_{\mathrm{An}(\{v\} \cup \mathrm{pa}(v) \cup \mathrm{nd}(v))}$ or from marriages.
- The two facts above imply that v is not connected directly to $\mathrm{nd}(v) \setminus \mathrm{pa}(v)$ in $\big(G_{\mathrm{An}(\{v\} \cup \mathrm{pa}(v) \cup \mathrm{nd}(v))}\big)^m$.

By (DG), $\quad A \perp B \mid S$ in $\big(G_{\mathrm{An}(A \cup B \cup S)}\big)^m \quad \Rightarrow \quad X_A \perp X_B \mid X_S$.

So

$$X_v \perp X_{\mathrm{nd}(v)} \mid X_{\mathrm{pa}(v)}. \tag{6}$$

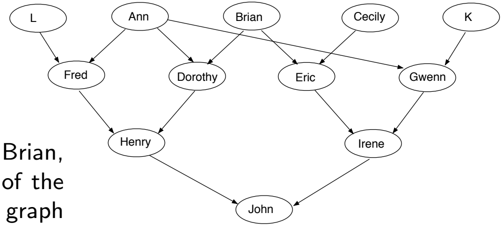
The property in (6) is called the *local directed Markov property* (DL). The discussion above shows also (DG) $\Rightarrow$ (DL).

Illustrations of (DL)

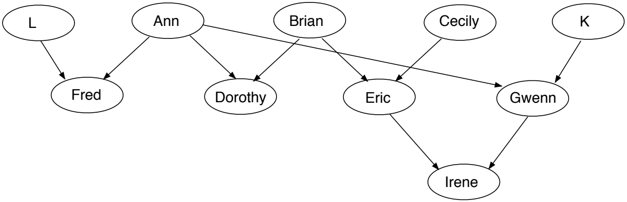
Example Stud-Farm:

$$\mathrm{pa}(\text{Dorothy}) = \{\text{Ann, Brian}\}$$

$$\mathrm{de}(\text{Dorothy}) = \{\text{Henry, John}\}$$



Given the genotypes of Ann and Brian, Dorothy's genotype is independent of the genotype of all the other nodes in the graph below.



Equivalence between Directed Markov Properties

In contrast to undirected graphical models, we have for a DAG G, the Markov properties (DF), (DG), (DL) and (DO) are all equivalent, where these properties are as defined earlier:

(DF) P factorizes recursively according to G, i.e.,

$$p(x) = \prod_{v \in V} p\big(x_v \,|\, x_{\mathrm{pa}(v)}\big).$$

(DG) P obeys the global directed Markov property, i.e.,

$$A \perp B \,|\, S \quad \text{in} \quad \big(G_{\mathrm{An}(A \cup B \cup S)}\big)^m \quad \Rightarrow \quad X_A \perp X_B \,|\, X_S.$$

(DL) P obeys the local directed Markov property, i.e.,

$$X_v \perp X_{\mathrm{nd}(v)} \mid X_{\mathrm{pa}(v)}.$$

(DO) P obeys the ordered directed Markov property, i.e., with respect to any well-ordering,

$$X_v \perp X_{\mathrm{pr}(v)} \mid X_{\mathrm{pa}(v)}.$$

We have verified "(DF) $\Rightarrow$ (DG) $\Rightarrow$ (DL)" and "(DO) $\Rightarrow$ (DF)." It is also straightforward to show "(DL) $\Rightarrow$ (DO)" (exercise).

Further Readings and Announcements

For directed Markov properties:

1. Robert G. Cowell et al. *Probabilistic Networks and Expert Systems*, Springer, 2007. Chap. 5.3.

This lecture is mainly based on [1]; some examples used are from Chaps. 2 and 3 of [2] below.

For d-separation, see [2] for an introduction, and [3] for deep analysis:

2. Finn V. Jensen. *An Introduction to Bayesian Networks*. UCL Press, 1996. Chap. 2.
3. Judea Pearl. *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, 1988. Chap. 3.

Exercise-related announcements:

- Written solutions need to be handed in, if you have done them, whether you come to the exercise meetings or not. Please submit the earlier solutions you've completed, if you forgot to do so.
- If for research, travel-related, or other proper reasons, you need extra time to finish a problem set, you may ask for an extension in advance.