This lecture covers the Factorization Property \((F)\), which expresses that the probability distribution of a graph can be separated into independent products of functions defined over complete subsets of the graph. We prove that \((F) \Rightarrow (G)\), and that under certain circumstances, \((F) \Leftrightarrow (G)\). We also cover some other things.

Note: This Course Note incorporates and subsumes Handout 5, which was passed out in class.

### 13.1 Review

In the last couple lectures, we covered five properties of conditional independence:

\[
\begin{align*}
(C1) \quad & X \perp Y \mid Z \quad \Rightarrow \quad Y \perp X \mid Z \\
(C2) \quad & X \perp Y \mid Z \text{ and } U = h(X) \quad \Rightarrow \quad U \perp Y \mid Z \\
(C3) \quad & X \perp Y \mid Z \text{ and } U = h(X) \quad \Rightarrow \quad X \perp Y \mid Z, U \\
(C4) \quad & X \perp Y \mid Z \text{ and } X \perp W \mid (Y, Z) \quad \Rightarrow \quad X \perp (W, Y) \mid Z
\end{align*}
\]

and, if \(P(X, Y, Z)\) is positive and continuous:

\[
(C5) \quad X \perp Y \mid Z \text{ and } X \perp Z \mid Y \quad \Rightarrow \quad X \perp (Y, Z)
\]

Note that the requirement for \((C5)\) is not an “iff.” Specifically, having a positive and continuous distribution is sufficient for \((C5)\) to hold, but continuity is not necessary. However, positivity alone is insufficient. The precise necessary conditions for \((C5)\) is an open problem. Lauritzen describes the condition as “no non-trivial logical relationship between \(Y\) and \(Z\).”

As an example of how \((C5)\) does not hold for non-continuous (and non-positive) distributions, consider the following distribution of discrete variables \(X\), \(Y\), and \(Z\):
\[ P(X = 0, Y = 0, Z = 0) = \frac{1}{2} \]
\[ P(X = 0, Y = 0, Z = 1) = 0 \]
\[ P(X = 0, Y = 1, Z = 0) = 0 \]
\[ P(X = 0, Y = 1, Z = 1) = 0 \]
\[ P(X = 1, Y = 0, Z = 0) = 0 \]
\[ P(X = 1, Y = 0, Z = 1) = 0 \]
\[ P(X = 1, Y = 1, Z = 0) = 0 \]
\[ P(X = 1, Y = 1, Z = 1) = \frac{1}{2} \]

Note that this is a corrected version of the example from last week, which assigned probabilities of 0.125 to the first and last cases.

Reasoning intuitively, \( X \perp Y \mid Z \) is true because knowing \( Z \) completely determines both \( X \) and \( Y \); hence they’re probabilistically independent. \( X \perp Z \mid Y \) is true by applying the same reasoning to \( Y, X, \) and \( Z \). However, \( X \not\perp (Y, Z) \) because knowing both \( Y \) and \( Z \) determines \( X \). This intuitive explanation can be quantitatively verified by factoring the joint probability distribution according to the definition of conditional independence.

Further recall the Markov properties for undirected graphs:

**(P) Pairwise** \( \alpha \perp \beta \mid V \setminus \{\alpha, \beta\} \) for non-adjacent \( \alpha, \beta \in V \)

**(L) Local** \( \alpha \perp V \setminus \text{cl}(\alpha) \mid \text{bd}(\alpha) \) for any \( \alpha \in V \)

**(G) Global** \( A \perp B \mid S \) for any disjoint subsets \( A \subset V, B \subset V, \) and \( S \subset V \) such that \( S \) separates \( A \) and \( B \)

We previously showed that:

\[ (G) \Rightarrow (L) \Rightarrow (P) \]

Furthermore, if for all disjoint subsets \( A, B, C, \) and \( D, \)

\[ A \perp B \mid C \cup D \text{ and } A \perp C \mid B \cup D \Rightarrow A \perp (B \cup C) \mid D \]

then

\[ (G) \Leftrightarrow (L) \Leftrightarrow (P) \]

### 13.2 Factorization Property

**Definition 13.1. Factorization**

A probability distribution \( P \) on \( X \) is said to factorize (F) according to \( G \) iff for all complete subsets \( a \subseteq V \), there exist nonnegative functions \( \psi_a \) that depend on \( X \) only through \( X_a \), and
The fact that the product ranges only over complete subsets of \( V \) is important. Any probability distribution can be factored into products of subsets that are not complete; simply factor it into power set of all subsets, and set \( \psi_{V'} \) to the distribution itself, and all other \( \psi_{\alpha} \) to 1.

Some examples:

Consider Graph \( G_1 \). Given any of the Markov properties, \( G_1 \) factorizes because each \( v_i \) is a clique, and independent of all other variables. Consequently, the covariance matrix is diagonal, and can be factored as the product of unary gaussians.

Consider Graph \( G_2 \). The cliques of this graph are \( \{v_1, v_2\} \), \( \{v_2, v_3\} \), etc. This graph factorizes if the probability distribution can be expressed as:

\[
P(X) = \psi_{v_1, v_2}(X) \psi_{v_2, v_3}(X) \psi_{v_3, v_4}(X) \psi_{v_4, v_5}(X)
\]

Consider Graph \( G_3 \). If the joint probability distribution \( P(v_1, v_2, v_3, v_4) \) can be factorized, then the most general factorization is:
This factorization includes all complete subsets. Often, however, the factorization appears simpler because some of the \( \psi \) are equal to 1.

The \( \psi \) may be defined arbitrarily. Therefore, when a factorization includes two products \( \psi_A \) and \( \psi_B \), and \( B \subset A \), then they can be represented by a new \( \psi_A' = \psi_A \psi_B \). If all \( \phi_B \) are removed in this way, such that the variables associated with any factor are not a subset of the variables associated any other factor, then the factorization is over all cliques (because the complete subsets are all maximal):

\[
P(X) = \prod_{c \in \mathcal{C}} \psi_c(X)
\]

where \( \mathcal{C} \) is the set of all cliques in \( G \). This factorization is unique.

**Theorem 13.2.** Factorization of a probability distribution \( P \) over an undirected graph \( \mathcal{G} \) implies the global Markov property \( (G) \).

That is:

\[
(F) \Rightarrow (G) \Rightarrow (L) \Rightarrow (P)
\]

**Proof.** We already know \( (G) \Rightarrow (L) \Rightarrow (P) \); we only need to show that \( (F) \Rightarrow (G) \).

Let \( (A, B, S) \) be any triple of disjoint subsets such that \( S \) separates \( A \) from \( B \). Also, let \( \tilde{A} \) be the connectivity components of \( \mathcal{G}_{V \setminus S} \) containing \( A \). I.e.,

\[
\tilde{A} = \bigcup_{\alpha \in A} [\alpha]_{V \setminus S}
\]

Figure 13.4: \( \tilde{A} \) contains the components connected to \( A \), and \( \tilde{B} \) the connected components of \( V \) not in \( \tilde{A} \) or \( S \)

Define \( \tilde{B} \) as \( \tilde{B} = V \setminus (\tilde{A} \cup S) \), so that \( \tilde{B} \) is the remainder of the graph. Note that \( B \subset \tilde{B} \).

Since \( A \) is separated from \( B \) by \( S \), \( A \) and \( B \) are in different connectivity components of \( \mathcal{G}_{V \setminus S} \). This implies that any clique of \( \mathcal{G} \) is either a subset of \( \tilde{A} \cup S \) or of \( \tilde{B} \cup S \) but not both (i.e., if we had a clique in both, some elements would bypass \( S \) and connect \( \tilde{A} \) to \( \tilde{B} \), but this cannot be since \( S \) separates \( A \) from \( B \)).
Let $\mathcal{C}_B$ be the cliques in $\tilde{A} \cup S$. Then we have that $\mathcal{C}$ (the cliques in the graph) are such that

$$\mathcal{C} = \mathcal{C}_A \cup \mathcal{C}_B$$

with $\mathcal{C}_B = \mathcal{C} \setminus \mathcal{C}_A$. But because of factorization, we can represent the joint distribution as:

$$f(x) = \prod_{c \in \mathcal{C}_A} \psi_c(x) = \prod_{c \in \mathcal{C}_A} \psi_c(x) \prod_{c \in \mathcal{C} \setminus \mathcal{C}_A} \psi_c(x) = h(x_{\tilde{A} \cup S}) h(x_{\tilde{B} \cup S})$$

which implies that $\tilde{A} \perp \tilde{B} \mid S$. But this implies that $A \perp B \mid S$ (the global Markov property (G)), because $A \subseteq \tilde{A}$ and $B \subseteq \tilde{B}$.

Note that this proof does not require positivity.

\[\square\]

### 13.3 The Hammersley and Clifford Theorem

**Theorem 13.3.** (Hammersley and Clifford) A probability distribution $P$ with positive and continuous density $f$ satisfies the pairwise Markov property with respect to an undirected graph $S$ if and only if it factorizes according to $S$. I.e., $(F) \equiv (G)$

Or in other words, if $P$ is both positive and continuous, then:

$$(F) \iff (G) \iff (L) \iff (P)$$

The proof is accomplished by showing that under these circumstances, $(P) \Rightarrow (F)$, thus completing the cycle of logical implication since we already know $(G) \Rightarrow (L) \Rightarrow (P)$.

Before beginning the proof, let us recall our notation. $X$ is the collection of random variables and $X_A$ is the subset of random variables for some set $A \subseteq V$ where $V$ is the set of nodes in the graph. An assignment to the set of random variables is denoted by $X = x$ or just $x$, and an assignment to a subset of the random variables is denoted as $X_A = x_A$ or just $x_A$.

As long as the probability density $f(x)$ is positive, it is possible to express the logarithm of a factored probability density as a sum:

$$\log f(x) = \sum_{\alpha \subseteq V} \phi_{\alpha}(x)$$

where $\phi_{\alpha}(x) = \log \psi_{\alpha}(x)$ and where we must have that $\phi_{\alpha}(x) \equiv 0$ unless $\alpha$ is a complete subset of $V$.

**Proof.** We shall proceed by defining appropriate $\psi_{\alpha}$ for every possible subset of $\alpha \subseteq V$, and then showing that the pairwise independence Markov property (P) implies that all $\psi_{\alpha}$ corresponding to non-complete subsets must be equal to one, and hence that there exists a valid factorization.

However, we must first prove the Möbius Inversion Lemma, which will be necessary for the proof. The Möbius inversion lemma will be useful because we will define a log-probability distribution composed of various $\phi_{\alpha}$ factors in the form of the first equation, and then apply the pairwise Markov independence property to the second equation.

Here we present an expanded version of the proof of the lemma that can be found in Lauritzen (p.239):
Lemma 13.4. (Möbius Inversion Lemma) Let \( \psi \) and \( \phi \) be functions defined on the set of all subsets of a finite set \( V \), taking values in an Abelian group (e.g., the reals). Then the following two equations are equivalent.

\[
\forall a \subseteq V : \psi(a) = \sum_{b \subseteq a} \phi(b) \tag{13.1}
\]

\[
\forall a \subseteq V : \phi(a) = \sum_{b \subseteq a} (-1)^{|a \setminus b|} \psi(b) \tag{13.2}
\]

Proof.

\[
\psi(a) = \sum_{b \subseteq a} \phi(b)
\]

\[
= \sum_{b \subseteq a} \sum_{c \subseteq b} (-1)^{|b \setminus c|} \psi(c)
\]

\[
= \sum_{c \subseteq a} \sum_{b \subseteq c \cap b \subseteq a} (-1)^{|b \setminus c|} \psi(c)
\]

\[
= \sum_{c \subseteq a} \psi(c) \sum_{h : c \subseteq b \cap h \subseteq a} (-1)^{|b \setminus c|}
\]

\[
= \sum_{c \subseteq a} \psi(c) \sum_{h : h \subseteq a \setminus c} (-1)^{|h|}
\]

The last step follows because the set of subsets \( b : c \subseteq b \subseteq a \) is like the set of subsets of \( a \setminus c \) since, by requiring \( c \subseteq b \), we are essentially reducing the number of possible subsets by a factor of \( 2^{|c|} \). That is, in each case there are a total of \( 2^{|a| - |c|} = 2^{|a \setminus c|} \) possible subsets, and the cardinalities of the set of subsets are the same. Also, for each of those subsets we are raising \( (-1)^{|c|} \) to the number of elements in that subset, but not including \( c \) (this is the exponent \(|b \setminus c|\)). An easy way to see this is to think of a bit vector to select subsets. In any event, the result then becomes the last equation.
Also, note that

$$\sum_{h : h \subseteq a \setminus c} (-1)^{|h|}$$

is zero for all $a \setminus c$ except for the case when $a \setminus c = \emptyset$ (i.e., for any non-empty set, there are the same number of even and odd subsets). Also, $a \setminus c = \emptyset$ only when $a = c$, leading to

$$\sum_{c : c \subseteq a} \psi(c) \sum_{h : h \subseteq a \setminus c} (-1)^{|h|} = \psi(a)$$

The lemma may also be proved in the other direction by substituting 13.1 into 13.2, with similar results.

We now return to the Hammersley and Clifford Theorem proper. Assume that $P$ is pairwise Markov. Choose a fixed but arbitrary assignment $x^*$ to the set of random variables $X$. For all $a \subseteq V$, define the following function

$$H_a(x) = \log f(x_a, x_{a^c}^*)$$

where $a^c \overset{\Delta}{=} V \setminus a$, and where we use the notation $(x_a, x_{a^c}^*)$ to indicate a particular assignment of values to all the random variables $X$. This assignment is defined as follows: for the random variables $X_a$, the assigned values come from the values contained in $x$, the argument of $H_a(x)$. For the random variables $X_{a^c} = X \setminus x_a$, the assigned values come from the arbitrary assignment we choose earlier, $x^*$. Another way of saying this is that $(x_a, x_{a^c}^*)$ is an assignment to all the random variables, which we denote by $\hat{x}$, such that $\hat{x}_\gamma = x^*_\gamma$ for $\gamma \notin a$ and where $\hat{x}_\gamma = x^*_\gamma$ for $\gamma \not\in a$.

Here’s an example that will make things clear if they are not already. Suppose $V = \{1, 2, 3, 4\}$ which means we have the four random variables $X = \{X_1, X_2, X_3, X_4\}$. We might choose some $x^*$ as an assignment. I.e., specifying $x^*$ could mean that $X_1 = 3, X_2 = 5, X_3 = 10, X_4 = 15$. Suppose that $a = \{1, 3\}$. Then the function $H_a(x)$ becomes

$$H_a(x) = H_a(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4) = \log f(x_1 = x_1, x_2 = 5, x_3 = x_3, x_4 = 15).$$

Note that $H_a(x)$ doesn’t depend on $x_2$ and $x_4$ since those are already assigned via $x^*$.

More generally, since $x^*$ is fixed, $H_a(x)$ depends on $x$ only via the values $x_a$ and not via $x_{a^c}$.

We further define for all $a \subseteq V$

$$\phi_a(x) = \sum_{b : b \subseteq a} (-1)^{|\alpha|} H_b(x)$$

Since $\phi_a(x)$ depends on $x$ only via $H_b(x)$ and since $b$ is chosen to be a subset of $a$, $\phi_a(x)$ also depends on $x$ only through $x_a$.

Next, apply the Möbius inversion lemma to obtain that

$$H_V(x) = \sum_{\alpha : \alpha \subseteq V} \phi_\alpha(x)$$

(note that the lemma says that this is also true for all $H_b, b \subseteq V$, but we need only $H_V$).

You might have noticed one important thing, namely that

$$\log f(x) = H_V(x)$$

because of the definition of $H_a(x)$. We therefore have that

$$\log f(x) = H_V(x) = \sum_{\alpha \subseteq V} \phi_\alpha(x)$$
This is the structure of the factorization. But to be a true factorization, we still must show that the summation contains $\phi_\alpha$ only if $a$ is a complete subset of $V$. We must ensure that any function $\phi_\alpha$ defined for a non-complete subset $a$ must be 0, thus ensuring that the corresponding $\psi_\alpha$ is 1.

So, assume that $\alpha, \beta \in a$ and that $\alpha \neq \beta$ (i.e., they are not joined, so $a$ is not complete). Also, define the set $c = a \setminus \{\alpha, \beta\}$. Then we can expand the definition of $\phi_\alpha(x)$ as follows:

$$
\phi_\alpha(x) = \sum_{b \subseteq a} (-1)^{|a \setminus b|} H_b(x)
$$

(expression of $a$ in terms of $c$)

$$
= \sum_{b \subseteq c} (-1)^{|a \setminus b|} H_b(x) + \sum_{b \subseteq c} (-1)^{|a \setminus (b \cup \{\alpha\})|} H_{b \cup \{\alpha\}}(x)
$$

(break up summation)

$$
= \sum_{b \subseteq c} (-1)^{|a \setminus b|} H_b(x) - \sum_{b \subseteq c} (-1)^{|a \setminus b|} H_{b \cup \{\alpha\}}(x)
$$

(simplify exponents)

$$
= \sum_{b \subseteq c} (-1)^{|a \setminus b|} (H_b(x) - H_{b \cup \{\alpha\}}(x) - H_{b \cup \{\beta\}}(x) + H_{b \cup \{\alpha, \beta\}}(x))
$$

(isolate summation factor)

$$
= \sum_{b \subseteq c} (-1)^{|c \setminus b|} (H_b(x) - H_{b \cup \{\alpha\}}(x) - H_{b \cup \{\beta\}}(x) + H_{b \cup \{\alpha, \beta\}}(x))
$$

(because $|a| = |c| + 2$)

So our job is done if we show that the quantity

$$
(H_b(x) - H_{b \cup \{\alpha\}}(x) - H_{b \cup \{\beta\}}(x) + H_{b \cup \{\alpha, \beta\}}(x))
$$

is zero. We do this as follows. First, for notational simplicity, define $d = V \setminus \{\alpha, \beta\}$. Then the following equations follow where we use positivity and continuity of the distributions:

$$
H_{b \cup \{\alpha, \beta\}}(x) - H_{b \cup \{\alpha\}}(x) = \log \frac{f(x_b, x_\alpha, x_\beta, x_{d \setminus \{\alpha, \beta\}})}{f(x_b, x_\alpha, x_\beta, x_{d \setminus \{\alpha, \beta\}})}
$$

(by the pairwise Markov property)

$$
= \log \frac{f(x_\alpha | x_\beta, x_b, x_{d \setminus \{\alpha, \beta\}})}{f(x_\alpha | x_\beta, x_b, x_{d \setminus \{\alpha, \beta\}})}
$$

(since the first ratio is just unity)

$$
= \log \frac{f(x_\alpha | x_\beta, x_b, x_{d \setminus \{\alpha, \beta\}})}{f(x_\alpha | x_\beta, x_b, x_{d \setminus \{\alpha, \beta\}})}
$$

(by pairwise Markov property and chain rule)

therefore everything is zero.

$\square$
As an example of how this theorem may be applied, consider the following distribution.

\[ Y \overset{\Delta}{=} (Y_1, Y_2, Y_3)^T \sim N(0, \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix}) \]

Then

\[ \Sigma_{13|2} = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{pmatrix} \Rightarrow Y_1 \perp Y_1 | Y_2 \]

\( Y \) satisfies (G) with respect to the graph in Figure 5.

![Figure 13.6:](image)

This distribution can therefore be factorized as

\[ f(Y_1, Y_2, Y_3) = \psi(Y_1, Y_2)\psi(Y_2, Y_3) \]

The concentration matrix \( K \) for this distribution is:

\[ K = \Sigma^{-1} = \begin{pmatrix} 2 & -2 & 0 \\ -2 & 6 & 2 \\ 0 & -2 & 2 \end{pmatrix} \]

Note that \( K_{13} = K_{31} = 0 \), thus indicating the global Markov property.

### 13.4 Relationship between Factorization and Decomposibility

Recall the definitions of decomposition and decomposable:

A *decomposition* \((A, B, S)\) of \( \mathcal{G} \) exists if \( A, B, \) and \( S \) are all disjoint subsets of the vertices \( V, V = A \cup B \cup S, S \) separates \( A \) and \( B, \) and \( S \) is complete.

A graph \( \mathcal{G} \) is *decomposable* if it is complete or if there exists a decomposition \((A, B, S)\) such that \( \mathcal{G}_{AUS} \) and \( \mathcal{G}_{BJS} \) are decomposable.

**Theorem 13.5.** If \((A, B, S)\) decomposes \( \mathcal{G} \), then \( P \) factorizes with respect to \( \mathcal{G} \) iff \( P_{AUS} \) and \( P_{BJS} \) factorize with respect to \( \mathcal{G}_{AUS} \) and \( \mathcal{G}_{BJS} \) and

\[ P(X) = \frac{P_{AUS}(X_{AUS})P_{BJS}(X_{BJS})}{P_S(X_S)} \]
Proof. (⇒) By the factorization property, we know that

\[ P(X) = \prod_{c \in \mathcal{G}} \psi_c(X) \]

Since \( S \) separates \( A \) and \( B \), there can be no cliques containing members in all of \( A, B, \) and \( S \); otherwise, there would be direct connections between the members in \( A \) and the members in \( B \). Hence, all cliques are either in \( \mathcal{G}_{A \cup S} \) or in \( \mathcal{G}_{B \cup S} \). Let \( A \) denote the cliques in \( A \cup S \), and \( B \) those in \( B \cup S \).

Then:

\[ P(X) = h(X_{A \cup S})k(X_{B \cup S}) \]

where

\[ h(X_{A \cup S}) = \prod_{c \in A} \psi_c(X) \]
\[ k(X_{B \cup S}) = \prod_{c \in B \setminus A} \psi_c(X) \]

Marginalizing out \( A \) and \( B \) from the distribution produces

\[ P(X_{A \cup S}) = \tilde{h}(X_{A \cup S})\tilde{k}(X_S) \]

where

\[ \tilde{k}(X_S) = \int k(X_{B \cup S})dX_B \]

and similarly

\[ P(X_{B \cup S}) = \tilde{h}(X_{B \cup S})\tilde{h}(X_S) \]

where

\[ \tilde{h}(X_S) = \int h(X_{A \cup S})dX_A \]

and finally

\[ P(X_S) = \int \tilde{h}(X_{A \cup S})k(X_{B \cup S})dX_A dX_B = \tilde{h}(X_S)\tilde{k}(X_S) \]

The right hand side immediately follows.

(⇐) The definition \( P(X) \) is a valid factorization; all cliques are also complete vertex subsets of the more general graph \( \mathcal{G} \). \( \square \)
13.5 Relationship between Factorization and Junction Trees

A junction tree induces a factorization. For example: the probability distribution of the following junction tree

![Junction Tree Diagram]

Figure 13.7:

can be decomposed as:

\[ P(X) = \frac{P(X_{ABC})P(X_{BCD})P(X_{DEF})}{P(X_{BC})P(X_{D})} \]

13.6 Markov Semantics of Directed Graphs

We previously discussed two algorithms for determining independence properties from directed graphs: Bayes-Ball and \(d\)-separation. We now show that a third mechanism, directed factorization, is also equivalent.

**Definition 13.6. Directed Factorization.**

A probability distribution \(P\) admits a *directed factorization* (DF) with respect to a DAG \(D\) if

\[ P(X) = \prod_{v \in V} P(X_v \mid X_{pa(v)}) \]

where \(pa(v)\) are the parents of \(v\).

**Lemma 13.7.** If \(P\) has (DF) with respect to \(D\), then it factorizes according to \(D^m\) (the moralized graph) and therefore obeys (G) on \(D^m\).

**Proof.** In the moralized graph \(D^m\), subsets \(\{v\} \cup pa(v)\) are complete. Define:

\[ \psi_{\{v\} \cup pa(v)}(X) \triangleq P(X_v \mid X_{pa(v)}) \]

Therefore, \(P\) factorizes according to complete sets, and therefore (F) \(\Rightarrow\) (G) for directed graphs.

\[ \square \]

(DF) can thus be seen to be a directed version of (G). (L) also has an equivalent for directed graphs:

\[ \alpha \perp V \setminus cl(\alpha) \mid bl(\alpha) \]
where $\text{bl}(\alpha)$ is the *Markov blanket*, defined as:

$$
\text{bl}(\alpha) \overset{\Delta}{=} \text{pa}(\alpha) \cup \text{ch}(\alpha) \cup \{w : \text{ch}(w) \cap \text{ch}(\alpha) \neq \emptyset \}
$$

Informally, the Markov blanket is $\alpha$’s parents, children, and the parents of $\alpha$’s children. It serves the same purpose as the boundary in undirected graphs: It delineates variable dependencies.

For example, all nodes in the graph below except $\alpha$ are in $\text{bl}(\alpha)$.

![Figure 13.8: Markov Blanket](image)

### 13.7 Directed global Markov property

Recall the definition of the ancestral set $A$ as a set for which $\text{bd}(\alpha) \subseteq A$ for all $\alpha \in A$. $\text{An}(A)$ is defined as the smallest ancestral set containing $A$.

**Theorem 13.8.** Let $P$ have (DF) over $D$. Then $A \perp B \mid S$ whenever $A$ and $B$ are separated by $S$ in $(D_{\text{An}(A \cup B \cup S)})^m$.

This independence property is called the directed global Markov property (DG). It is in fact equivalent to directed factorization (DF).

Here’s an example:
Is \( a \perp b \mid \{X, Y\} \) in the graph at left? Yes, because in the corresponding moralized graph shown to the right, vertex \( Y \in S \) blocks the path between \( a \) and \( b \), thus \( S \) separates \( a \) and \( b \), and the independence holds.

\( d \)-separation can also be interpreted this way. A trail \( \pi \) from \( a \) to \( b \) is blocked in a directed graph \( D \) by \( S \) if \( \pi \) contains a node \( \gamma \in \pi \) such that either a) \( \gamma \in S \) and \( \rightarrow \gamma \rightarrow \rightarrow \) or \( \leftarrow \gamma \rightarrow \rightarrow \), or b) \( \gamma \notin S \) and \( \rightarrow \gamma \leftarrow \rightarrow \) and no descendants of \( \gamma \) are in \( S \). \( A \) and \( B \) are \( d \)-separated if all trails from \( A \) to \( B \) are blocked.