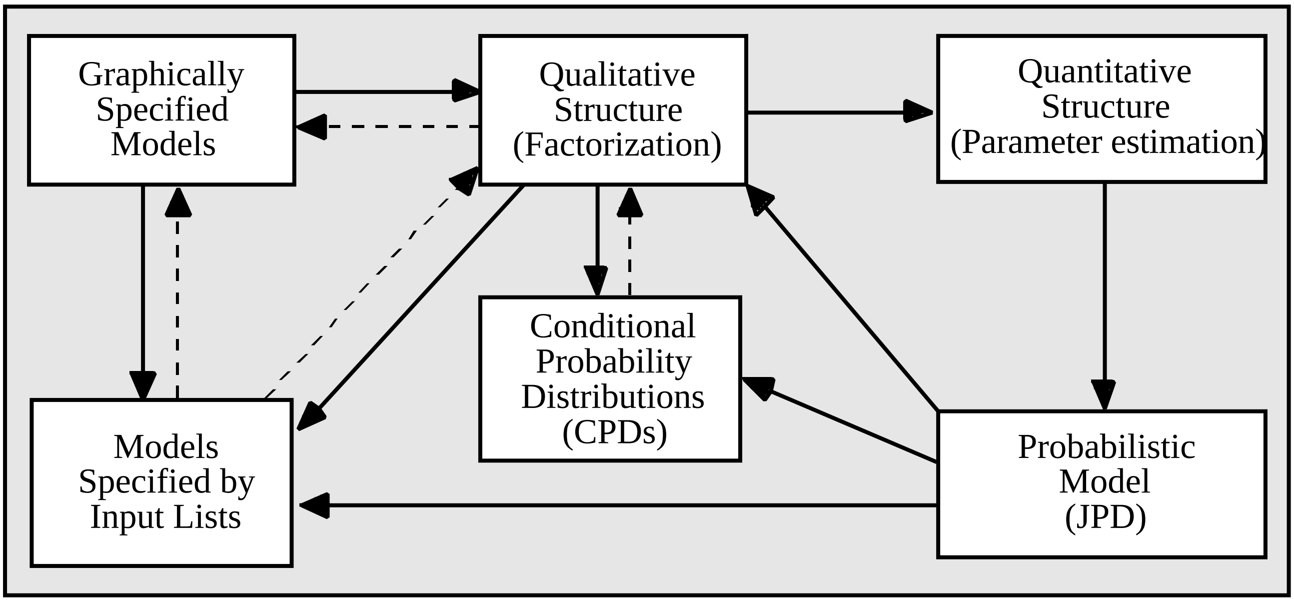
DEFINING PROBABILISTIC MODELS

The Joint Probability Distribution (JPD) of a set of n binary variables involve a huge number of parameters 2^n (larger than 10^{25} for only 100 variables).

x	y	z	$p(x, y, z)$
0	0	0	0.12
0	0	1	0.18
0	1	0	0.04
0	1	1	0.16
1	0	0	0.09
1	0	1	0.21
1	1	0	0.02
1	1	1	0.18

We can use the qualitative structure of the model to simplify the probabilistic structure.



Cantabria University

CONDITIONAL PROBABILITY. INDEPENDENCE

Definición 1 Conditional probability. *Let X and Y be two disjoint subsets of variables such that $p(y) > 0$. Then, the conditional probability distribution (CPD) of X given $Y = y$ is given by*

$$p(X = x|Y = y) = p(x|y) = \frac{p(x, y)}{p(y)}. \tag{1}$$

Definición 2 Independence of two variables. *Let X and Y be two disjoint subsets of the set of random variables $\{X_1, \ldots, X_n\}$. Then X is said to be independent of Y if and only if*

$$p(x|y) = p(x), \tag{2}$$

for all possible values x and y of X and Y; otherwise X is said to be dependent on Y.

Also, if X is independent of Y, we can then combine (1) and (2) and obtain which implies

$$p(x, y) = p(x)p(y). \tag{3}$$

If $\{X_1, \ldots, X_m\}$ are independent, then

$$p(x_1, \ldots, x_m) = \prod_{i=1}^{m} p(x_i), \tag{4}$$

Definición 3 Conditional independence. *Let X, Y and Z be three disjoint sets of variables, then X is said to be conditionally independent of Y given Z, if and only if*

$$p(x|z,y) = p(x|z). \quad \leftrightarrow \quad p(x,y|z) = p(x|z)p(y|z).$$

When X and Y are conditionally independent given Z, we write $I(X,Y|Z)$. The statement $I(X,Y|Z)$ is referred to as a **conditional independence statement** (CIS). Similarly, when X and Y are conditionally dependent given Z, we write $D(X,Y|Z)$, which is called a **conditional dependence statement**. The definition of conditional independence conveys the idea that once Z is known, knowing Y can no longer influence the probability of X. In other words, if Z is already known, knowledge of Y does not add any new information about X.

Note that (unconditional) independence can be treated as a particular case of conditional independence. For example, we can write $I(X,Y|\phi)$, to mean that X and Y are unconditionally independent.

Definición 4 Factorization by potentials. *Let* $C_1, \ldots, C_m$ *be subsets of a set of variables* $X = \{X_1, \ldots, X_n\}$.

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{m} \Psi_i(c_i),$$

where the functions Ψ_i*, called factor potentials, are nonnegative.*

Definición 5 Chain rule factorizations. *Any JPD of a set of ordered variables* $\{X_1, \ldots, X_n\}$ *can be expressed as a product of* m *CPDs of the form*

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{m} p(y_i|b_i), \tag{5}$$

where $B_i = \{Y_1, \ldots, Y_{i-1}\}$.

Ejemplo 1 Chain rule. Consider a case of four variables $\{X_1, \ldots, X_4\}$. Then the following are equivalent chain rule factorizations of the JPD:

$$p(x_1, \ldots, x_4) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2)p(x_4|x_1, x_2, x_3)$$

and

$$p(x_1, \ldots, x_4) = p(x_1|x_2, x_3, x_4)p(x_2|x_3, x_4)p(x_3|x_4)p(x_4).$$

Cantabria University

IMPOSING INDEPENDENCIES

Consider the variables $\{X_1, X_2, X_3, X_4\}$ and suppose we have:

$$I(X_3, X_1|X_2) \text{ and } I(X_4, \{X_1, X_3\}|X_2). \quad (6)$$

We wish to compute the constraints among the parameters of the JPD imposed by these CISs. The first of these statements implies

$$p(x_3|x_1, x_2) = p(x_3|x_2), \quad (7)$$

and the second statement implies

$$p(x_4|x_1, x_2, x_3) = p(x_4|x_2). \quad (8)$$

Note that the general form of the JPD is not a suitable representation for calculating the constraints given by (7) and (8). However, by using these two equalities we obtain

$$p(x_1, \ldots, x_4) = p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_2). \quad (9)$$

Therefore, the two CISs in (6) give rise to a reduction in the number of parameters from 15 to 7.

DEPENDENCY MODELS. SEPARATION

Definición 6 Dependency model. *Any model M of a set of variables $\{X_1, \ldots, X_n\}$ from which we can determine whether $I(X, Y|Z)$ is true, for all possible triplets of disjoint subsets X, Y, and Z, is called a dependency model.*

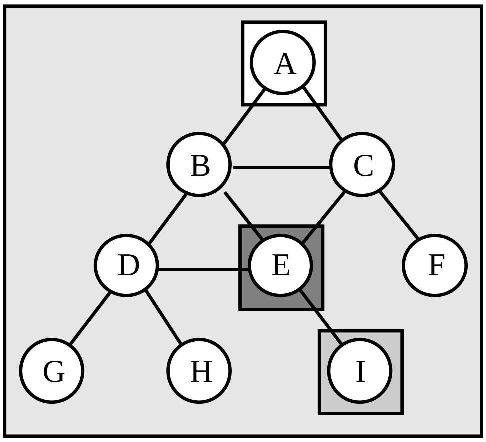
A JPD (with the definition of conditional independence) is a dependency model. A graph can also define a dependency model (using the corresponding separation criterion),
The qualitative structure of a probabilistic model can be represented by a graphical dependency model that provides a way to factorize the corresponding JPD.

Definición 7 U-separation. *Let X, Y, and Z be three disjoint subsets of nodes in an undirected graph G. We say that Z separates X and Y iff every path between each node in X and each node in Y contains at least one node in Z. When Z separates X and Y in G, we write $I(X, Y|Z)_G$; otherwise $D(X, Y|Z)_G$.*

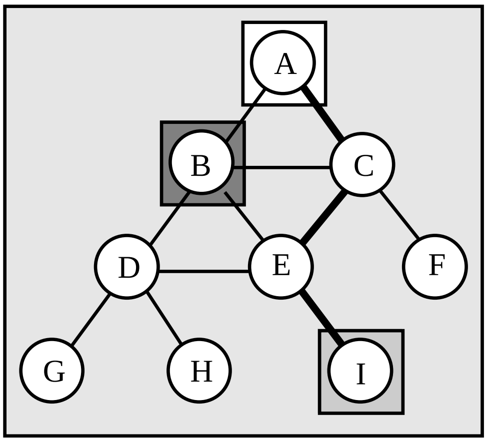
Given an undirected graph, one can derive all CISs from the graph using the above U-separation criterion.

Cantabria
University

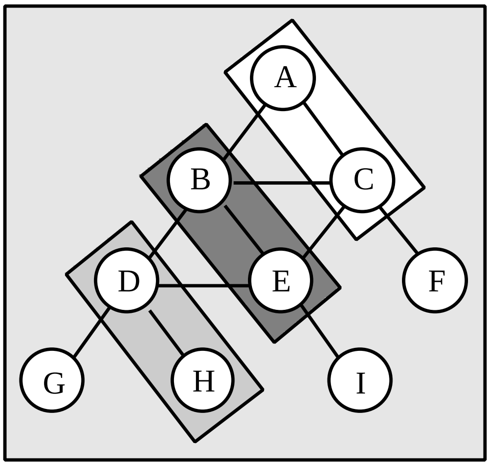
U-SEPARATION EXAMPLE



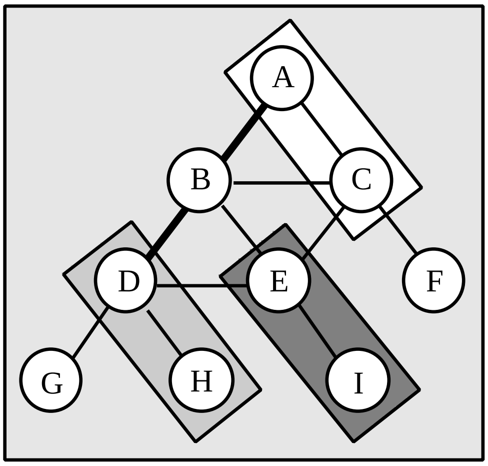
(a) $I(A, I \mid E)$



(b) $D(A, I \mid B)$



(c) $I(\{A, C\}, \{D, H\} \mid \{B, E\})$



(d) $D(\{A, C\}, \{D, H\} \mid \{E, I\})$

- Every path between A and I contains E. Thus, $I(A, I|E)_G$.
- There is a path $(A - C - E - I)$ that does not contain B. Thus $I(A, I|B)_G$.
- Every path between the two subsets contains either B or E. $I(\{A, C\}, \{D, H\}|\{B, E\})_G$.
- There is a path $(A - B - D\}$ that does not contain the variables E and I.

Definición 8 D-Separation. *Let X, Y, and Z be three disjoint subsets of nodes in a DAG D; then Z is said to D-separate X and Y, iff along every undirected path from each node in X to each node in Y there is an intermediate node A such that either*

1. *A is a head-to-head node in the path, and neither A nor its descendants are in Z, or*

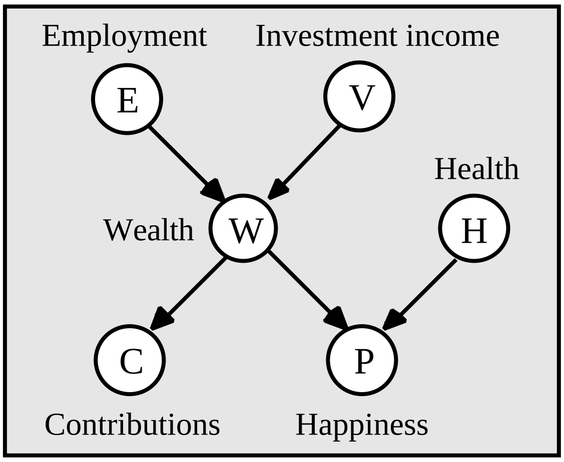
2. *A is not a head-to-head node in the path and A is in Z.*

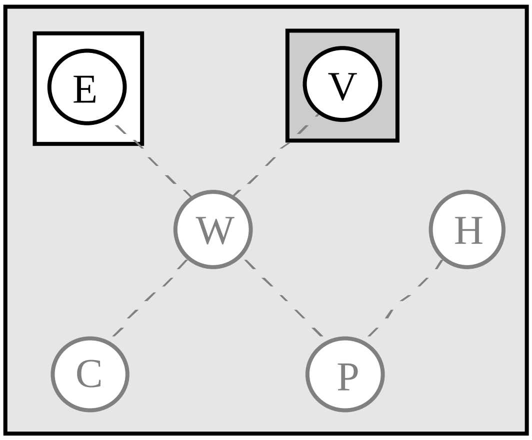
When Z D-separates X and Y in D, we write $I(X, Y|Z)_D$ to indicate that this CIS is derived from D; otherwise we write $D(X, Y|Z)_D$ to indicate that X and Y are conditionally dependent given Z in the graph D.

Definición 9 D-Separation. *Let X, Y, and Z be three disjoint subsets of nodes in a DAG D, then Z is said to D-separate X and Y iff Z separates X and Y in the moral graph of the smallest ancestral set containing X, Y, and Z.*

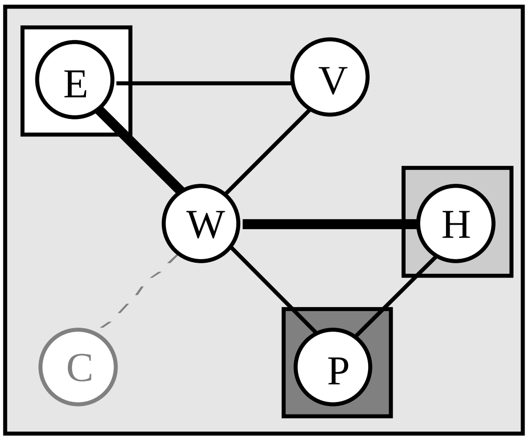
Cantabria
University

D-SEPARATION EXAMPLE

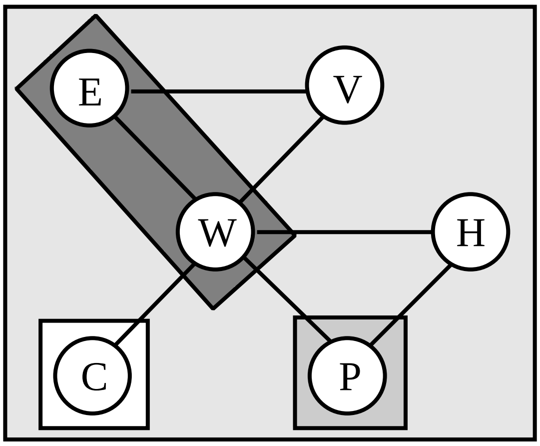




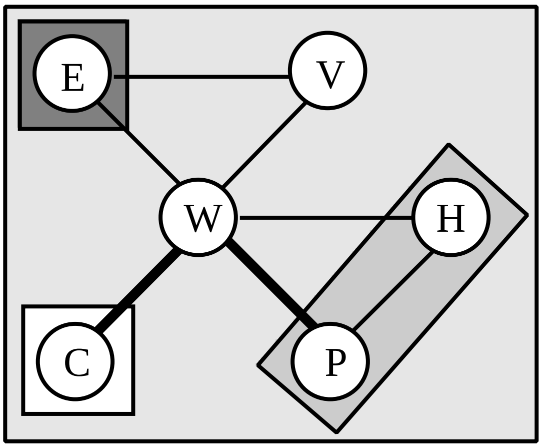
(a) $I(E, V \mid \emptyset)$



(b) $D(E, H \mid P)$



(c) $I(C, P \mid \{E, W\})$



(d) $D(C, \{H, P\} \mid E)$

Cantabria University

PROPERTIES OF CONDITIONAL INDEPENDCE

- **Symmetry:** If X is conditional independent (c.i.) of Y, then Y is c.i. of X

$$I(X, Y|Z) \Leftrightarrow I(Y, X|Z).$$

- **Decomposition:** If X is c.i. of $Y \cup W$ given Z, then X is c.i. of Y given Z, and X is c.i. of W given Z, that is,

$$I(X, Y \cup W|Z) \Rightarrow I(X, Y|Z) \text{ and } I(X, W|Z).$$

- **Weak Union:**

$$I(X, Y \cup W|Z) \Rightarrow I(X, W|Z \cup Y) \text{ and } I(X, Y|Z \cup W).$$

- **Contraction:** If W is irrelevant to X after the learning of some irrelevant information Y, then W must have been irrelevant before we knew Y, that is,

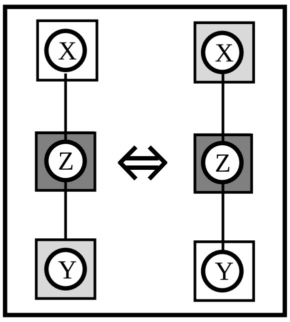
$$I(X, W|Z \cup Y) \; and \; I(X, Y|Z) \Rightarrow I(X, Y \cup W|Z).$$

The *weak union* and *contraction* properties together mean that irrelevant information should not alter the relevance of other relevant information in the system. In other words, what was relevant remains relevant, and what was irrelevant remains irrelevant.

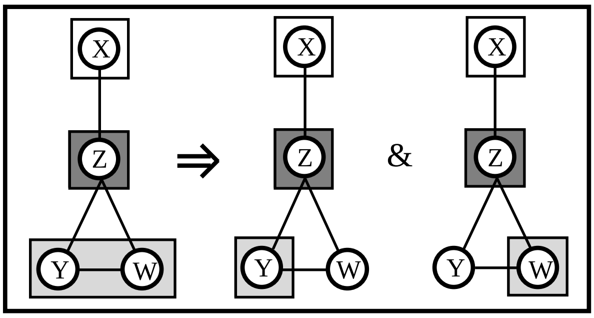
The above four properties hold for any JPD.

Cantabria University

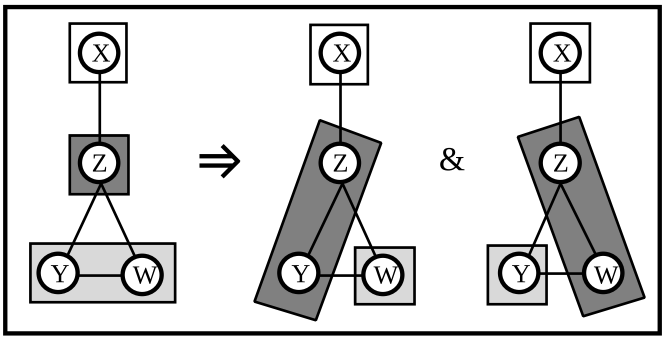
GRAPHICAL ILLUSTRATION



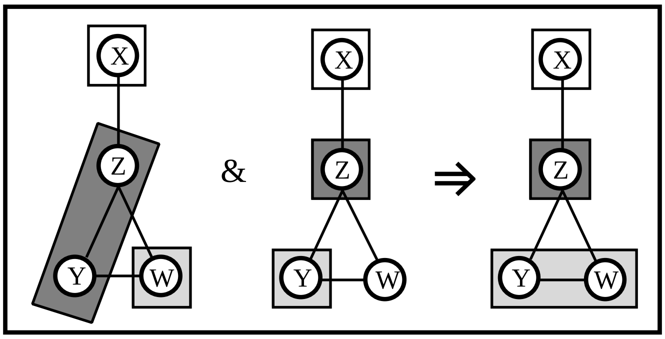
(a) Symmetry



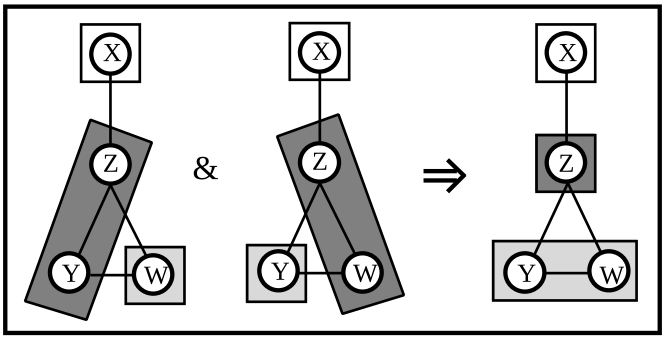
(b) Decomposition



(c) Weak Union



(d) Contraction



(e) Intersection

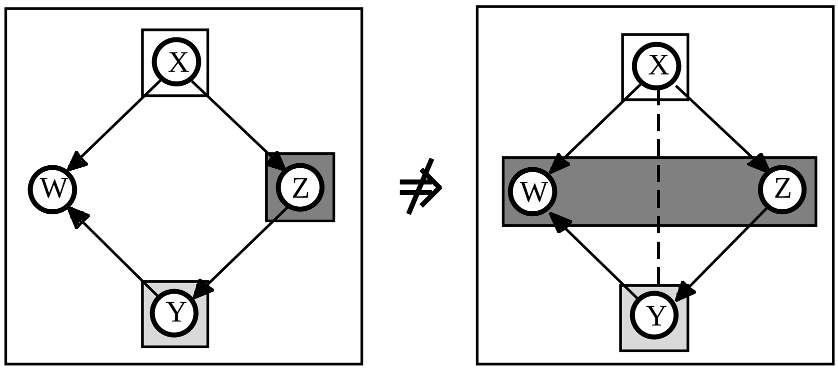
Cantabria
University

OTHER PROPERTIES

1. **Strong Union:** If X is c.i. of Y given Z, then X is also c.i. of Y given $Z \cup W$, that is,

$$I(X, Y|Z) \Rightarrow I(X, Y|Z \cup W).$$

Strong union property violation by DAGs.



(a) (b)

2. **Intersection:**

$$I(X, W|Z \cup Y) \text{ and } I(X, Y|Z \cup W) \Rightarrow I(X, Y \cup W|Z).$$

3. **Strong Transitivity:**

$$D(X, A|Z) \text{ and } D(A, Y|Z) \Rightarrow D(X, Y|Z),$$

4. **Weak Transitivity:**

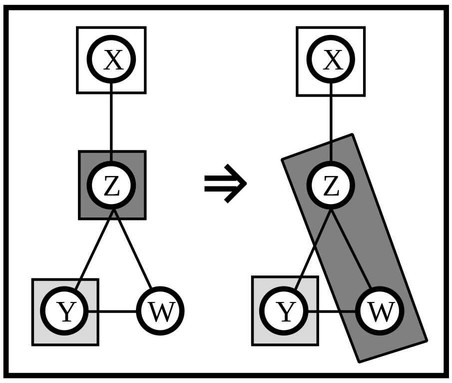
$$D(X, A|Z) \text{ and } D(A, Y|Z) \Rightarrow D(X, Y|Z) \text{ or } D(X, Y|Z \cup A),$$

5. **Chordality:**

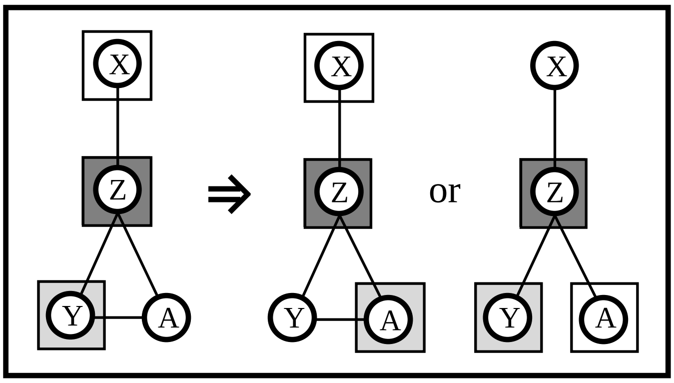
$$D(A, C|B) \text{ and } D(A, C|D) \Rightarrow D(A, C|B \cup D) \text{ or } D(B, D|A \cup C),$$

Cantabria
University

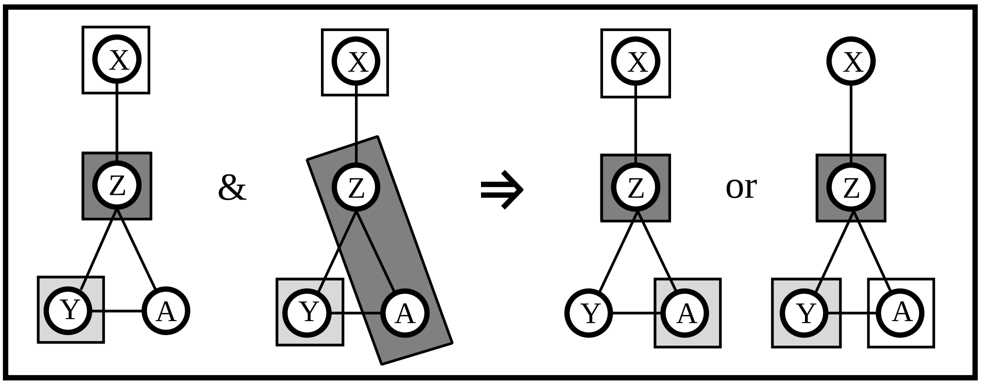
GRAPHICAL ILLUSTRATION



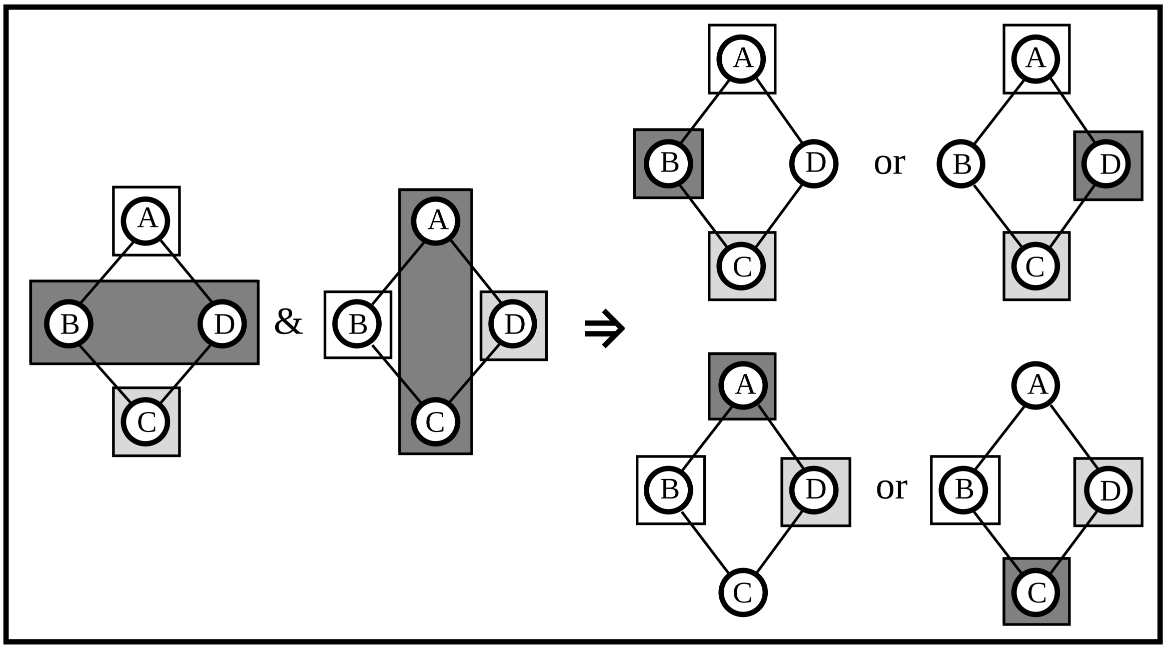
(a) Strong Union



(b) Strong Transitivity



(c) Weak Transitivity



(d) Chordality

Graphs display the relationships among the variables explicitly and are intuitive and easy to explain. It is important to analyze whether or not dependence models associated with probabilistic models can be given by graphical models.

Definición 10 Perfect map. *A graph G is said to be a perfect map of a dependency model M if every CIS derived from G can also be derived from M and vice versa, that is,*

$$I(X,Y|Z)_M \Leftrightarrow I(X,Y|Z)_G \Leftrightarrow Z \textit{ separates } X \textit{ from } Y.$$

Unfortunately, not every dependency model can be represented by a directed or undirected perfect map.

Ejemplo 2 Dependency model with no directed perfect map. *Consider the set of three variables $\{X,Y,Z\}$ and the dependency model*

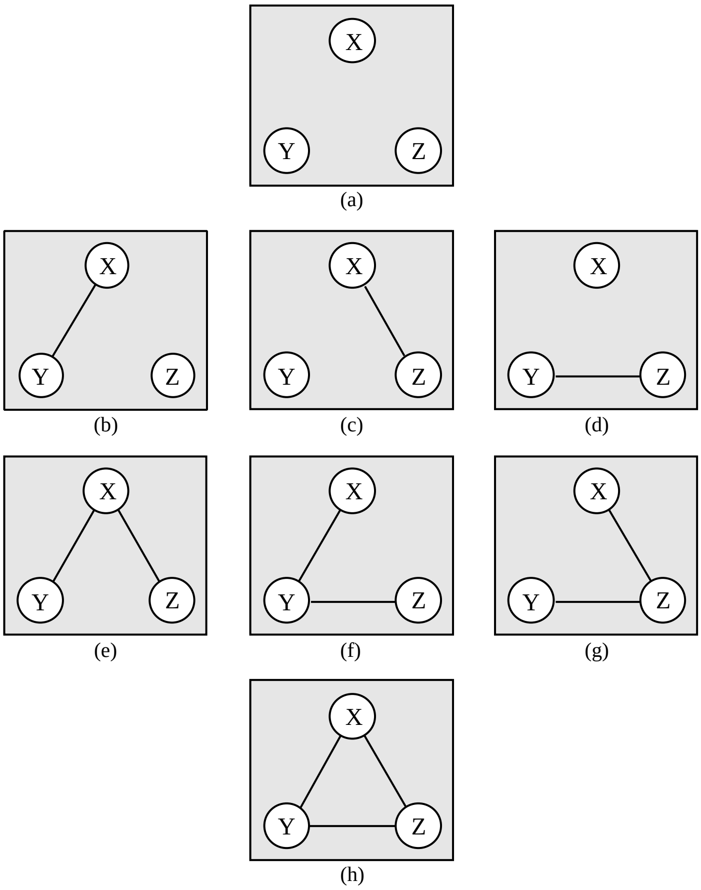
$$M = \{I(X,Y|Z), I(Y,Z|X), I(Y,X|Z), I(Z,Y|X)\}.$$

There is no directed acyclic graph (DAG) D that is a perfect map of the dependency model M.

Cantabria
University

MODEL WITH NO UNDIRECTED PERFECT MAP

$$M = \{I(X, Y|\phi), I(Y, X|\phi)\},$$



Graph G	G not in M	M not in G
(a)	$I(X, Z\|\phi)$	ϕ
(b)	$I(X, Z\|\phi)$	$I(X, Y\|\phi)$
(c)	$I(Y, Z\|\phi)$	ϕ
(d)	$I(X, Z\|\phi)$	ϕ
(e)	$I(Y, Z\|X)$	$I(X, Y\|\phi)$
(f)	$I(X, Z\|Y)$	$I(X, Y\|\phi)$
(g)	$I(X, Y\|Z)$	$I(X, Y\|\phi)$
(h)	ϕ	$I(X, Y\|\phi)$

Cantabria
University

I-MAPS AND D-MAPS

Definición 11 Independence map. *A graph G is said to be an independence map (I-map) of a dependency model M if*

$$I(X,Y|Z)_G \Rightarrow I(X,Y|Z)_M,$$

that is, if all CISs derived from G hold in M.

Definición 12 Dependency map. *A graph G is said to be a dependency map (D-map) of a dependency model M if*

$$D(X,Y|Z)_G \Rightarrow D(X,Y|Z)_M,$$

that is, all CISs derived from G hold in M.

Definición 13 Minimal I-map. *A graph G is said to be a minimal I-map of a dependency model M if it is an I-map of M, but it is not an I-map of M when removing any link from it.*

Cantabria University

MODELS WITH UNDIRECTED PERFECT MAPS

A necessary and sufficient condition for a dependency model M to have an undirected perfect map is that M must satisfy the following properties:

- **Symmetry:**

$$I(X,Y|Z)_M \Leftrightarrow I(Y,X|Z)_M.$$

- **Decomposition:**

$$I(X,Y \cup W|Z)_M \Rightarrow I(X,Y|Z)_M \text{ and } I(X,W|Z)_M.$$

- **Intersection:**

$$I(X,W|Z \cup Y)_M \text{ and } I(X,Y|Z \cup W)_M \Rightarrow I(X,Y \cup W|Z)_M.$$

- **Strong union:**

$$I(X,Y|Z)_M \Rightarrow I(X,Y|Z \cup W)_M.$$

- **Strong transitivity:**

$$I(X,Y|Z)_M \Rightarrow I(X,A|Z)_M \text{ or } I(Y,A|Z)_M,$$

where A is a single node not in $\{X,Y,Z\}$.

Definición 14 Markov network. *A Markov network is a pair* (G, Ψ) *where* G *is an undirected graph and* $\Psi = \{\psi_1(c_1), \ldots, \psi_m(c_m)\}$ *is a set of positive potential functions defined on the cliques* $C_1, \ldots, C_m$ *of* G *that defines the JPD* $p(x)$ *as*

$$p(x) = \prod_{i=1}^{n} \psi_i(c_i). \tag{10}$$

If the undirected graph G *is triangulated, then* $p(x)$ *can also be factorized, using probability functions* $P = \{p(r_1|s_1), \ldots, p(r_m|s_m)\}$*, as*

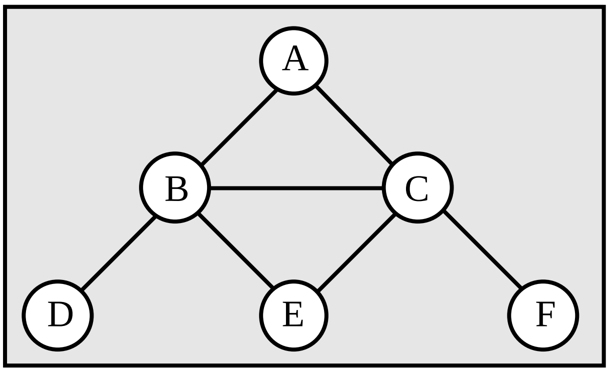
$$p(x_1, \ldots, x_n) = \prod_{i=1}^{m} p(r_i|s_i), \tag{11}$$

where R_i *and* S_i *are the separator and residual of the cliques. In this case, the Markov network model is defined by* (G, P)*. The graph* G *is an undirected* I*-map of* $p(x)$*.*

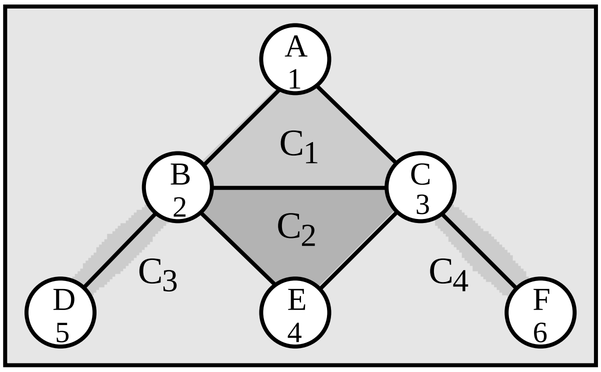
Thus, a Markov network can be used to define the qualitative structure of a probabilistic model through a factorization of the corresponding JPD in terms of potential functions or probability functions. The quantitative structure is then obtained by numerically specifying the functions appearing in the factorization.

Cantabria
University

MARKOV NETWORKS
EXAMPLE



(a)



(b)

The cliques of the graph are :

$$\begin{array}{ll} C_1 = \{A, B, C\}, & C_2 = \{B, C, E\}, \\ C_3 = \{B, D\}, & C_4 = \{C, F\}. \end{array} \tag{12}$$

$$\begin{aligned} p(a, b, c, d, e, f) &= \psi_1(c_1)\psi_2(c_2)\psi_3(c_3)\psi_4(c_4) \\ &= \psi_1(a, b, c)\psi_2(b, c, e)\psi_3(b, d)\psi_4(c, f). \end{aligned} \tag{13}$$

Since the graph is triangulated, another factorization of the JPD in terms of probability functions can be obtained

i	Clique C_i	Separator S_i	Residual R_i
1	A, B, C	ϕ	A, B, C
2	B, C, E	B, C	E
3	B, D	B	D
4	C, F	C	F

$$\begin{aligned} p(a, b, c, d, e, f) &= \prod_{i=1}^{4} p(r_i|s_i) \\ &= p(a, b, c)p(e|b, c)p(d|b)p(f|c). \end{aligned} \tag{14}$$

Cantabria University

MODELS WITH DIRECTED PERFECT MAPS

A necessary condition for a dependency model M to have a directed perfect map is that M must satisfy the following properties:

- **Symmetry:** $I(X, Y|Z)_M \Leftrightarrow I(Y, X|Z)_M$.
- **Composition-Decomposition:**

$$I(X, Y \cup W|Z)_M \Leftrightarrow I(X, Y|Z)_M \text{ and } I(X, W|Z)_M.$$

- **Intersection:**

$$I(X, W|Z \cup Y)_M \text{ and } I(X, Y|Z \cup W)_M \Rightarrow I(X, Y \cup W|Z)_M.$$

- **Weak union:**

$$I(X, Y \cup Z|W)_M \Rightarrow I(X, Y|W \cup Z)_M.$$

- **Weak transitivity:**

$$I(X, Y|Z)_M \text{ and } I(X, Y|Z \cup A)_M \Rightarrow I(X, A|Z)_M \; or \, I(Y, A|Z)_M,$$

- **Contraction:**

$$I(X, Y|Z \cup W)_M \text{ and } I(X, W|Z)_M \Rightarrow I(X, Y \cup W|Z)_M.$$

- **Chordality:**

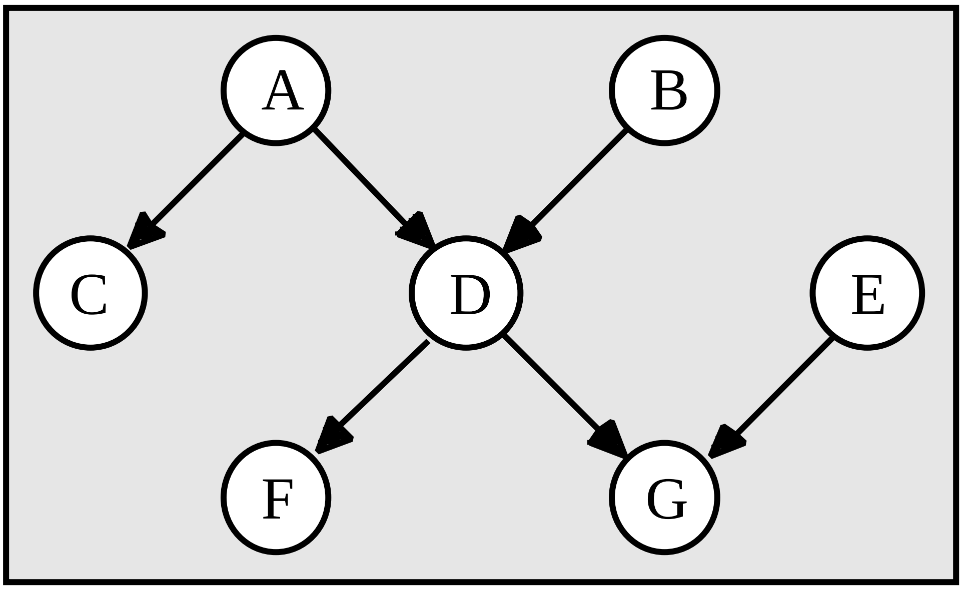
$$I(A, B|C \cup D)_M \, , \; I(C, D|A \cup B)_M \Rightarrow I(A, B|C)_M \text{ or } I(A, B|D)_M,$$

Definición 15 Bayesian network. *A Bayesian network is a pair (D, P), where D is a DAG, $P = \{p(x_1|\pi_1), \ldots, p(x_n|\pi_n)\}$ is a set of n CPDs, one for each variable, and Π_i is the set of parents of node X_i in D. The set P defines the associated JPD as*

$$p(x) = \prod_{i=1}^{n} p(x_i|\pi_i). \tag{15}$$

The DAG D is a minimal directed I-map of $p(x)$.

Example



$$p(X) = p(a)p(b)p(c|a)p(d|a, b)p(e)p(f|d)p(g|d, e).$$