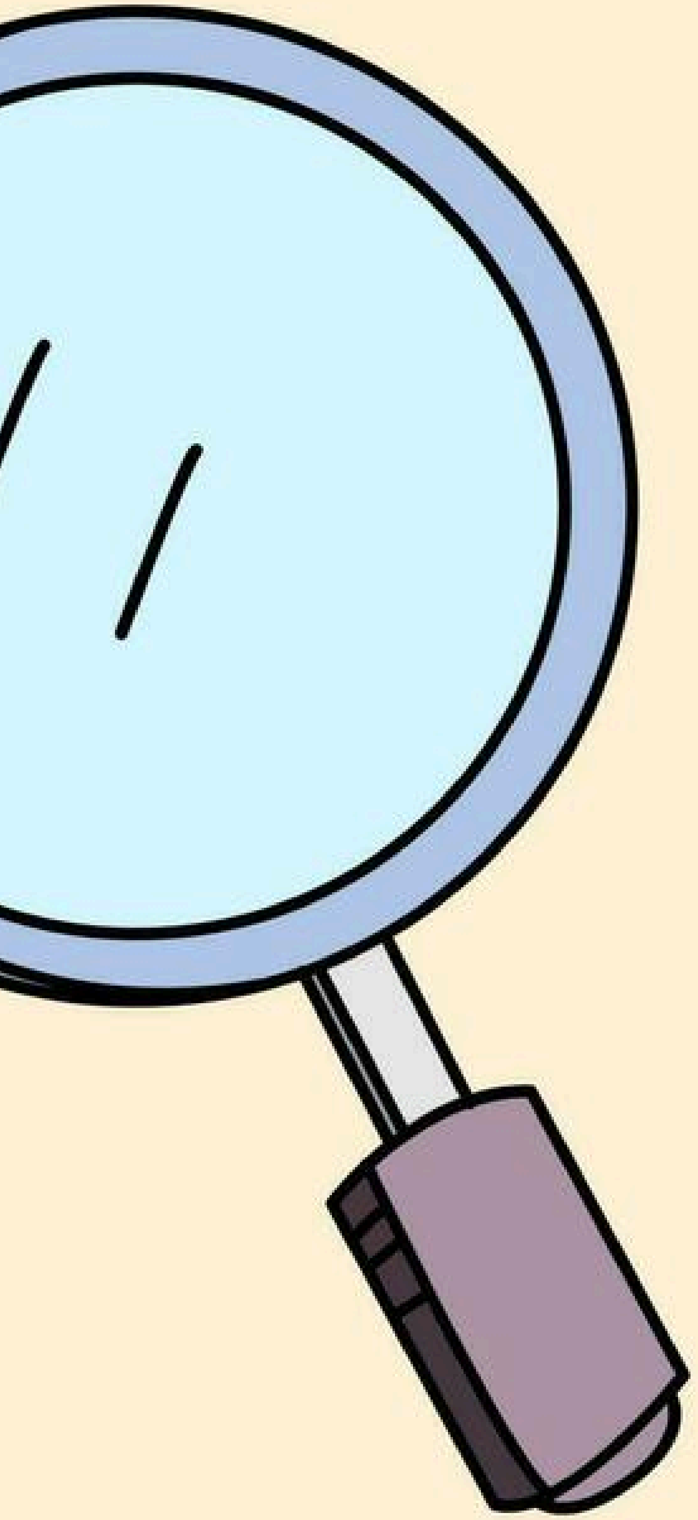
MINERÍA DE DATOS

Maximiliano Ojeda

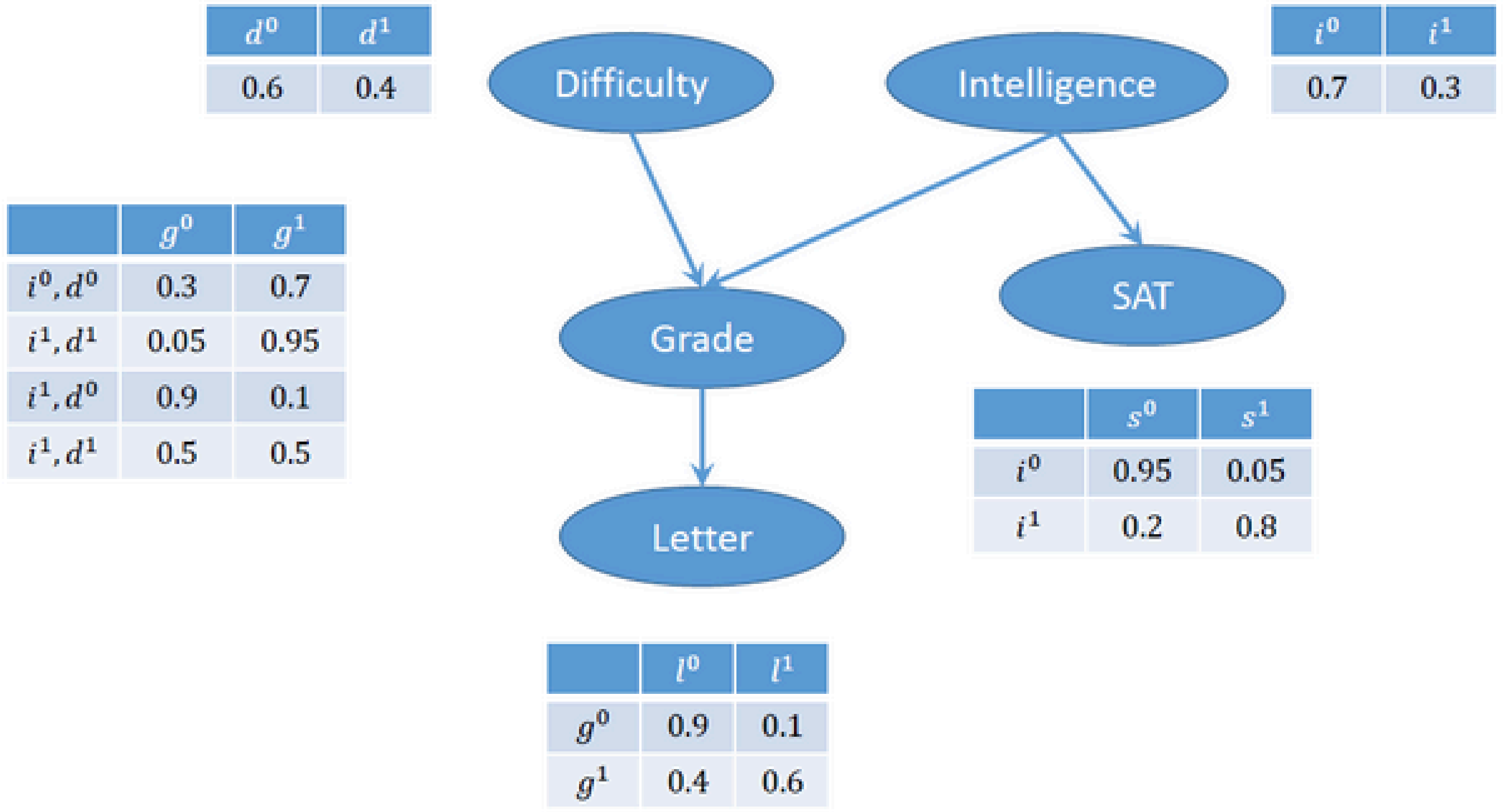
muojeda@uc.cl

IIC-2433



Redes Bayesianas

Redes Bayesianas



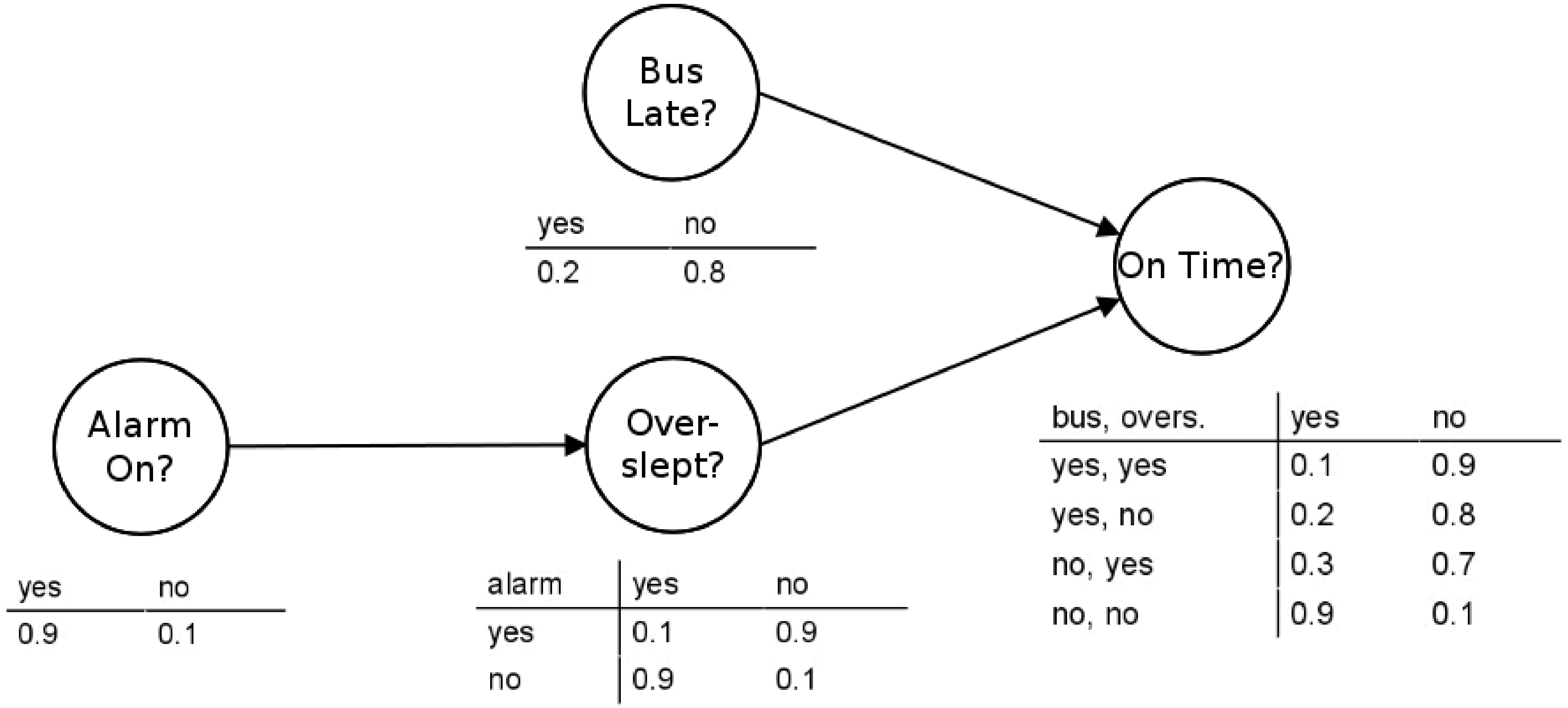
d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^0	g^1
i^0, d^0	0.3	0.7
i^1, d^1	0.05	0.95
i^1, d^0	0.9	0.1
i^1, d^1	0.5	0.5

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^0	0.9	0.1
g^1	0.4	0.6



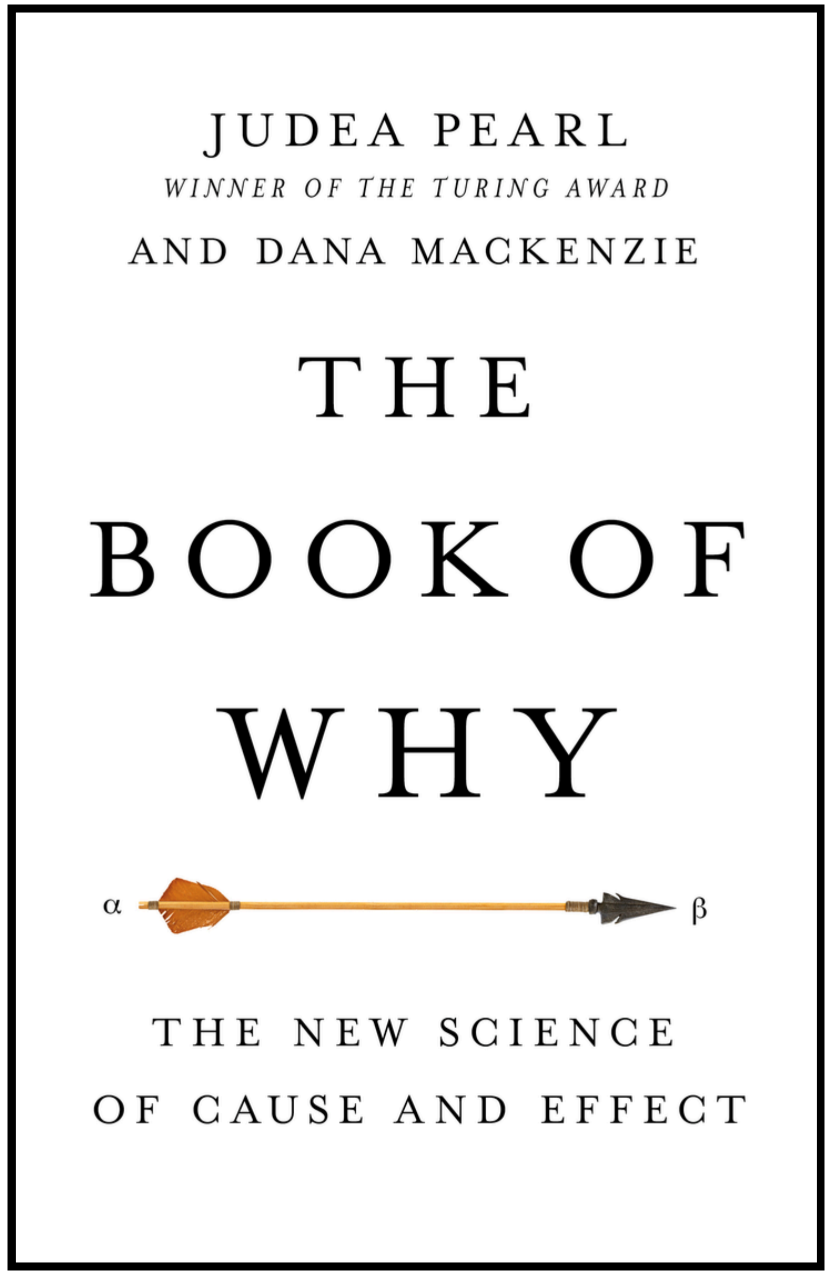
yes	no
0.2	0.8

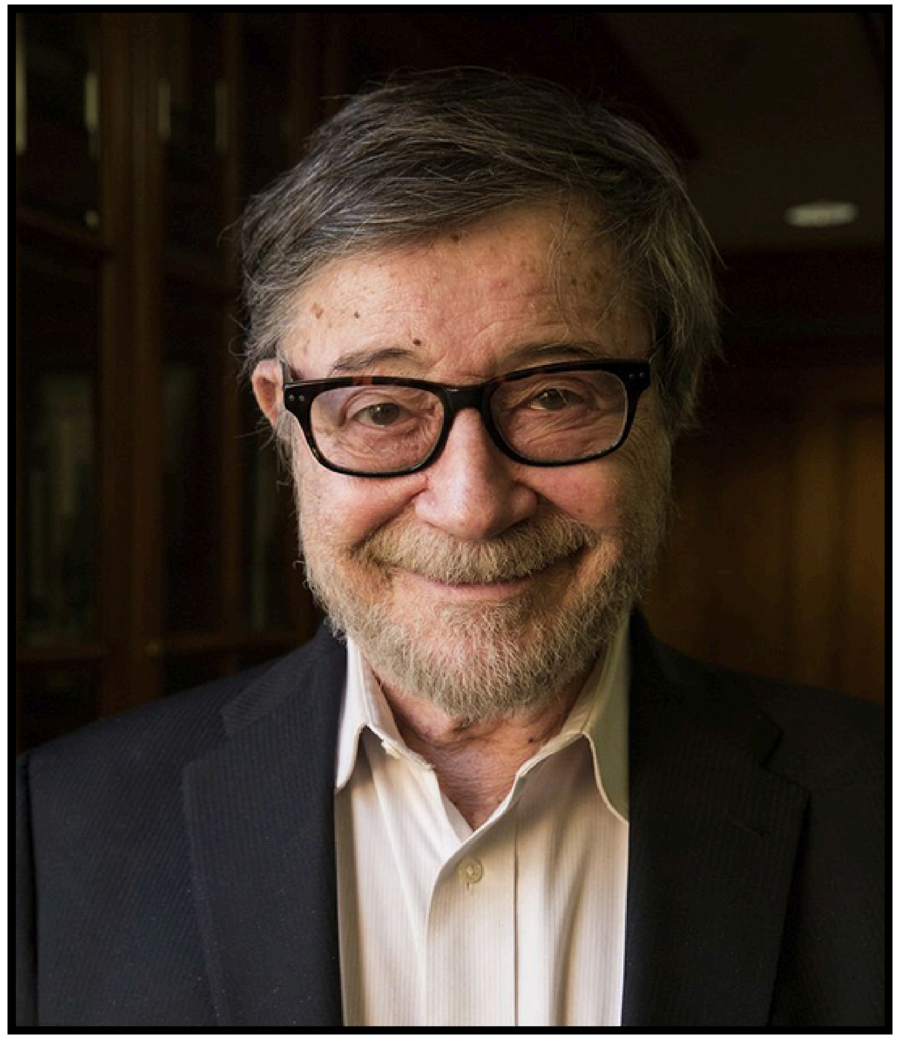
yes	no
0.9	0.1

alarm	yes	no
yes	0.1	0.9
no	0.9	0.1

bus, overs.	yes	no
yes, yes	0.1	0.9
yes, no	0.2	0.8
no, yes	0.3	0.7
no, no	0.9	0.1

Redes Bayesianas





Fundamentos de probabilidad

Antes de entrar en redes bayesianas, es necesario recordar cómo funcionan las probabilidades simples y condicionales.

Evento y variable aleatoria

Ej: X=“Tener fiebre”, puede ser “sí” o “no”.

Probabilidad conjunta

P(A,B) → probabilidad de que ocurran A y B al mismo tiempo.

Probabilidad condicional

P(A | B) → probabilidad de A dado que B ocurrió.
Se calcula como:

$$P(A \mid B) = \frac{P(A, B)}{P(B)}$$

Regla del producto

P(A | B) → probabilidad de A dado que B ocurrió

$$P(A, B) = P(A \mid B) \cdot P(B)$$

Teorema de Bayes

El Teorema de Bayes es famoso porque es la herramienta central para razonar bajo **incertidumbre**.

Sirve para actualizar creencias

$$P(\mathrm{A}|\mathrm{B}) = \frac{P(\mathrm{B}|\mathrm{A}) \cdot P(\mathrm{A})}{P(\mathrm{B})}$$

Thomas Bayes (1701 - 1761)

Independencia y dependencia condicional

Las redes bayesianas se basan en qué variables dependen o no dependen entre sí.

Independencia

$P(A, B) = P(A) \cdot P(B)$

Dependencia Condicional

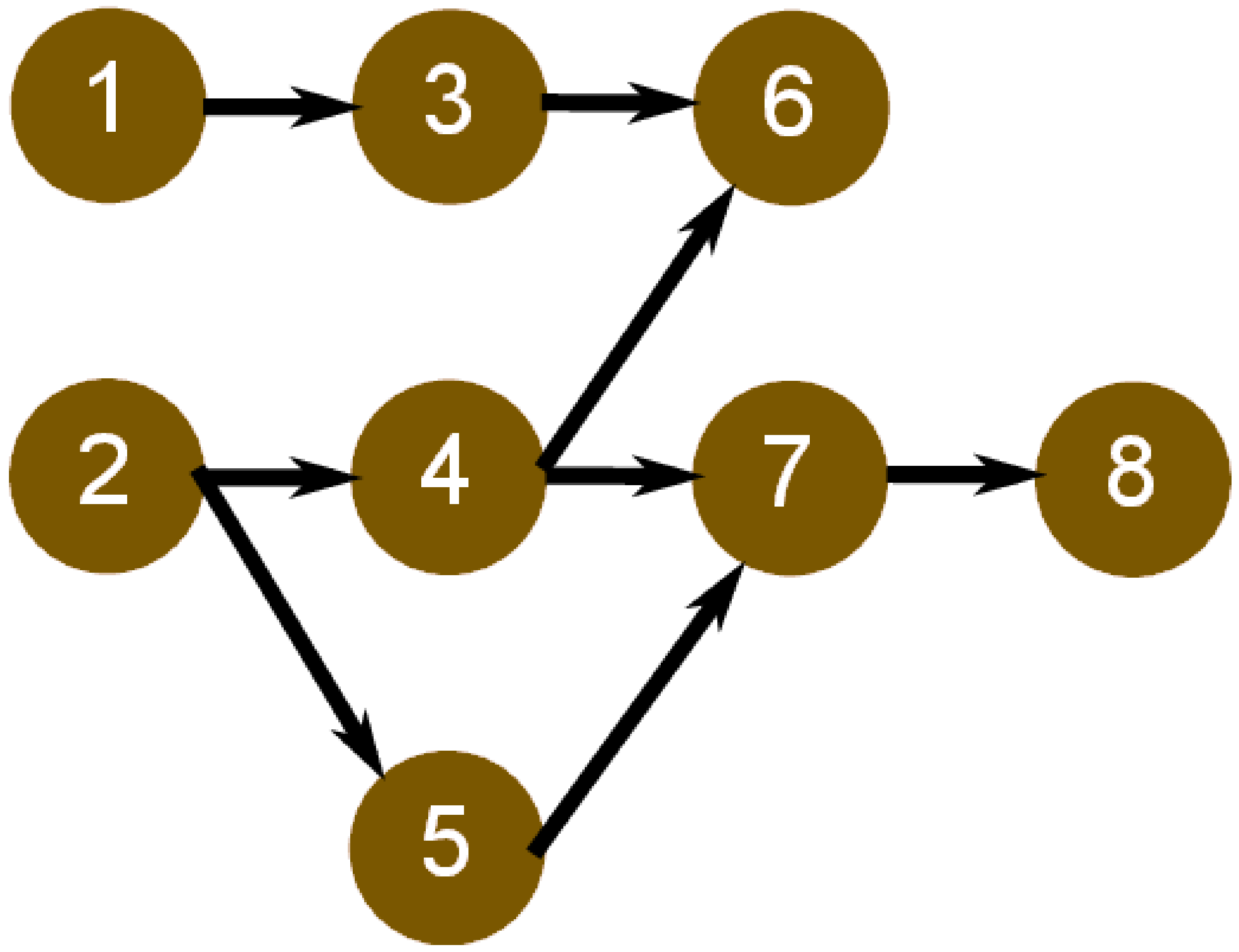
$P(A \mid B, C) = P(A \mid C)$ si A es independiente de B dado C

Esto se traduce directamente a la estructura del grafo:
si no hay flecha entre dos nodos, puede significar que son condicionalmente independientes.

Grafos dirigidos acíclicos (DAGs)

Es un tipo de grafo con dos características clave:
Dirigido: Cada arista tiene una dirección específica, apuntando desde un nodo hacia otro.

Acíclico: No contiene "ciclos". Esto significa que es imposible empezar en un nodo, seguir las flechas en su dirección, y terminar de nuevo en el nodo original.



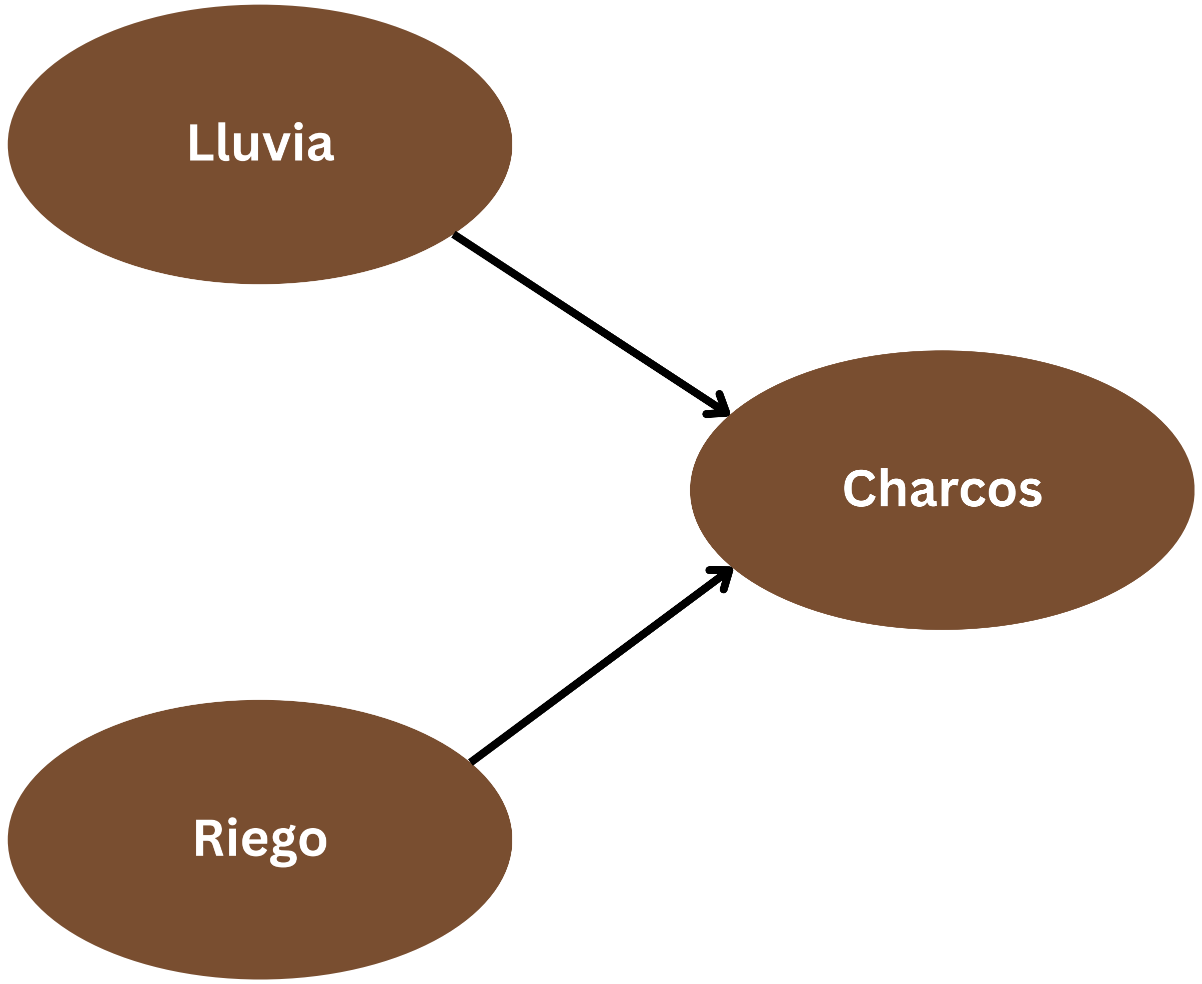
Una **red bayesiana** se representa con un grafo dirigido sin ciclos:

- Nodos → variables aleatorias
- Arcos (flechas) → dependencias o relaciones causales
- No hay ciclos (no puedes volver al mismo nodo)

Tablas de Probabilidad Condicional (CPT)

Cada nodo tiene su tabla de probabilidades condicionales

Lluvia	Charcos	P(Charcos | Lluvia)
Sí	Sí	0.9
Sí	No	0.1
No	Sí	0.2
No	No	0.8

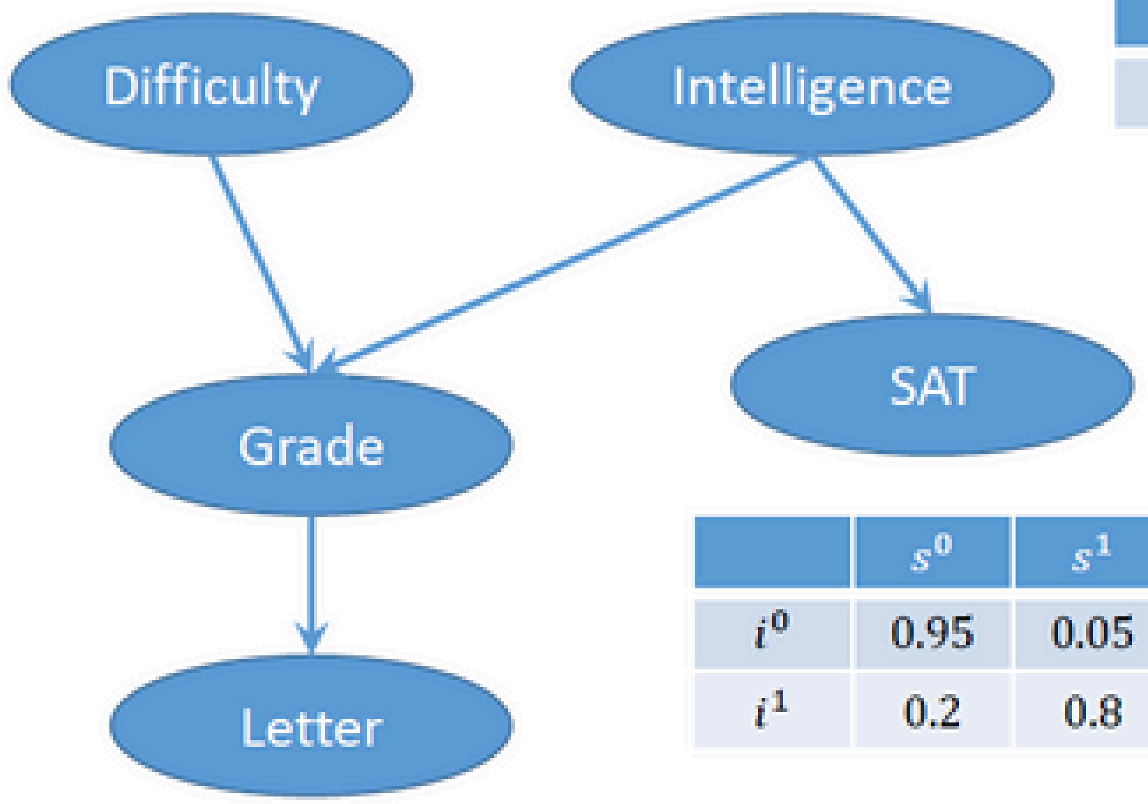


Ejemplo Student Bayesian Network

La red intenta modelar la probabilidad de que un estudiante obtenga una buena nota y una buena recomendación en función de su inteligencia y la dificultad del curso.

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3



	g^0	g^1
i^0, d^0	0.3	0.7
i^1, d^1	0.05	0.95
i^1, d^0	0.9	0.1
i^1, d^1	0.5	0.5

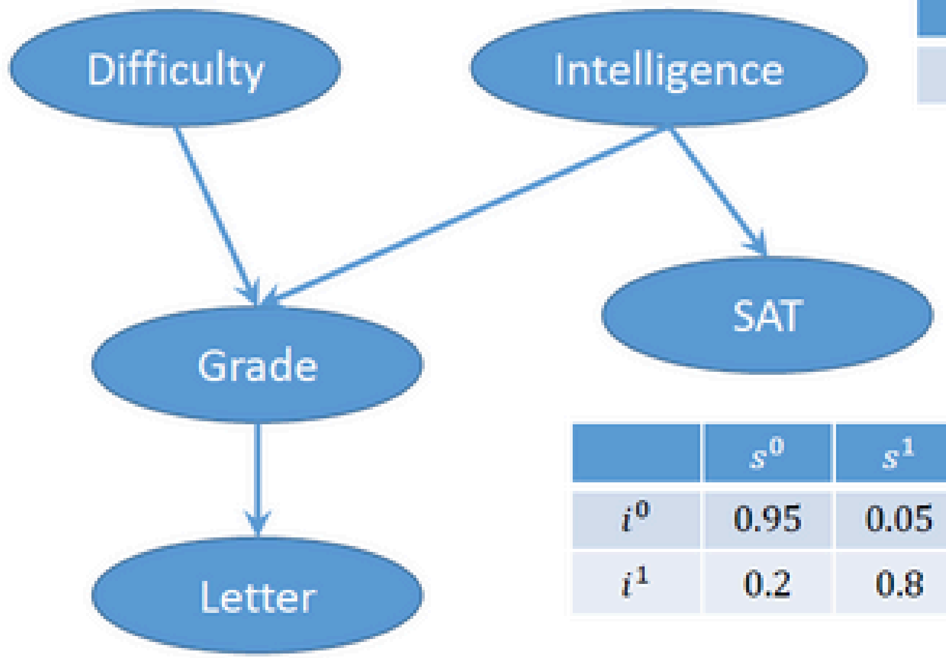
	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^0	0.9	0.1
g^1	0.4	0.6

Inferencia Student Bayesian Network

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3



	g^0	g^1
i^0, d^0	0.3	0.7
i^1, d^1	0.05	0.95
i^1, d^0	0.9	0.1
i^1, d^1	0.5	0.5

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^0	0.9	0.1
g^1	0.4	0.6

Las tablas no contienen toda la información

Ya hay muchas probabilidades, pero no hay directamente cosas como:

- P(Intelligence | Grade)
- P(Letter | Intelligence)
- P(SAT | Letter)
- P(Intelligence | SAT,Letter)

Inferencia Student Bayesian Network

La inferencia bayesiana significa actualizar creencias cuando existe evidencia parcial.

Ejemplo:

- No se sabe si el curso es difícil ni si el estudiante es inteligente.
- Obtuvo buena nota (Grade = bueno).

“¿Cuál es la probabilidad de que sea inteligente dado que obtuvo buena nota?”

Para calcularla, es necesario:

- P(Intelligence)
- P(Difficulty)
- P(Grade | Intelligence, Difficulty)

$$P(\text{Intelligence} \mid \text{Grade}) = \frac{P(\text{Grade} \mid \text{Intelligence})\, P(\text{Intelligence})}{P(\text{Grade})}$$

Inferencia Student Bayesian Network

Esto es lo que se conoce a priori

$P(\text{Intelligence} = \text{Alta}) = 0.3$

$P(\text{Difficulty} = \text{Dificil}) = 0.4$

$P(\text{Grade} = \text{Buena} \mid \text{Intelligence} = \text{Alta}, \text{Difficulty} = \text{Dificil}) = 0.5$

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^0	g^1
i^0, d^0	0.3	0.7
i^1, d^1	0.05	0.95
i^1, d^0	0.9	0.1
i^1, d^1	0.5	0.5

Queremos calcular esto

$$\boxed{P(\text{Intelligence} \mid \text{Grade}) = \frac{P(\text{Grade} \mid \text{Intelligence})\, P(\text{Intelligence})}{P(\text{Grade})}}$$

Solo necesitamos calcular probabilidades marginal

$$P(\text{Grade} = \text{Buena} \mid \text{Intelligence} = \text{Alta}) = \sum_{d \in \{\text{facil}, \text{dificil}\}} P(\text{Grade} = \text{Buena} \mid \text{Intelligence} = \text{Alta}, \text{Difficulty} = d)\, P(\text{Difficulty} = d)$$

$$P(\text{Grade} = \text{Buena}) = \sum_{i \in \{\text{baja}, \text{alta}\}} \sum_{d \in \{\text{facil}, \text{dificil}\}} P(\text{Grade} = \text{Buena} \mid i, d)\, P(i)\, P(d)$$

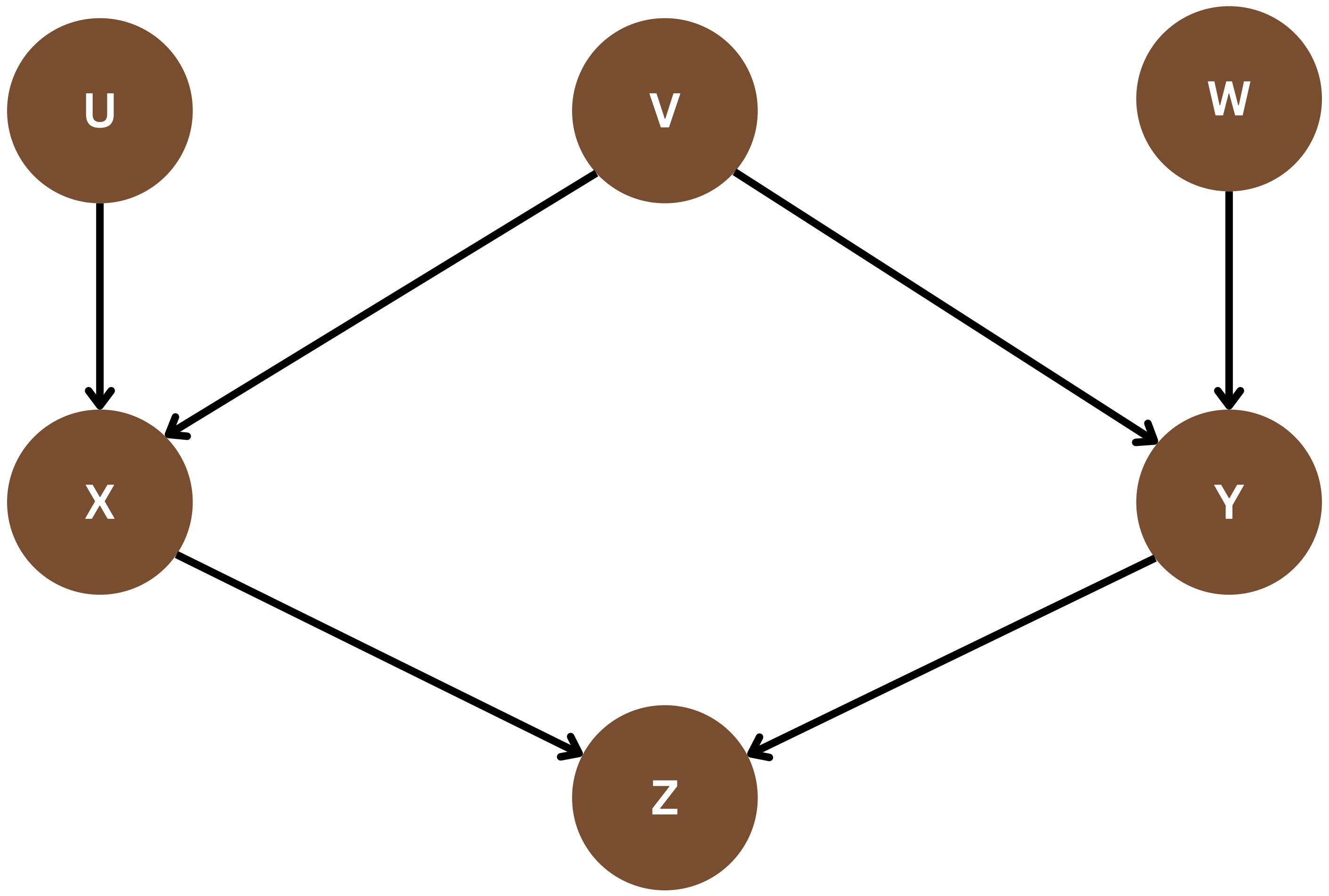
Inferencia: Variable Elimination

Es uno de los algoritmos centrales de **inferencia exacta** en Redes Bayesianas. Sirve para calcular **probabilidades marginales o condicionales** cuando la red tiene muchas variables conectadas.

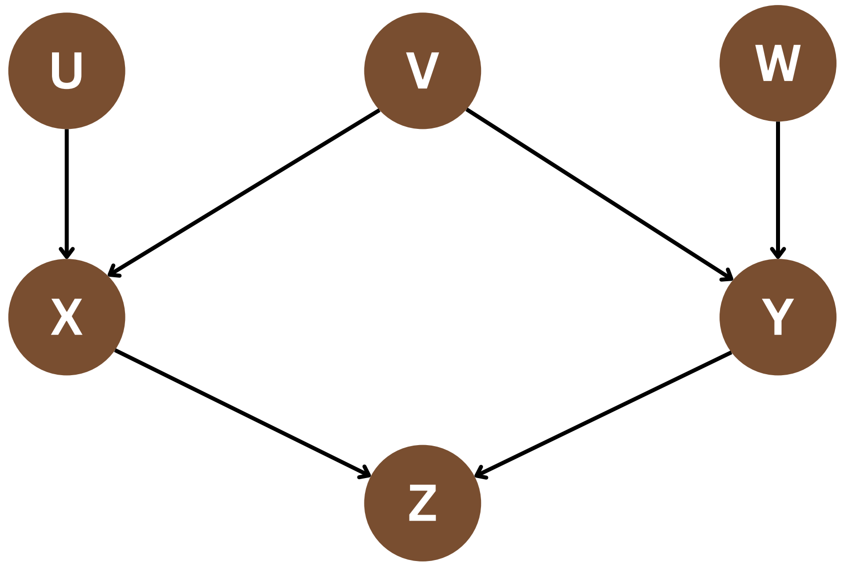
En lugar de sumar sobre todas las combinaciones posibles, VE "elimina" variables una a una, combinando solo los factores necesarios en cada paso.

Cada eliminación consiste en:

- Multiplicar los factores que contienen la variable.
- Sumar los valores posibles de esa variable (marginalizarla).
- Guardar el resultado como un nuevo factor reducido.



Inferencia: Variable Elimination



Query

$$P(U \mid +z)$$

Sin eliminación de variables

$$P(U \mid +z) \propto P(U) \sum_{v,w,x,y} P(v)\, P(w)\, P(x \mid U, v)\, P(y \mid v, w)\, P(+z \mid x, y)$$

Factores Iniciales

$$P(U); P(V); P(W); P(X \mid U, V); P(Y \mid V, W); P(+z \mid X, Y)$$

Orden de Eliminación

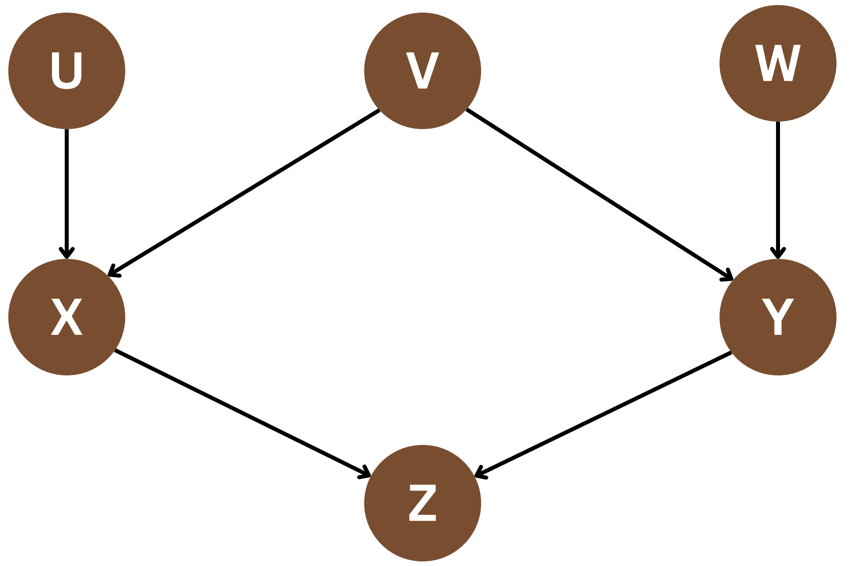
$$W, Y, V, X$$

Inferencia: Variable Elimination

Query $P(U \mid +z)$

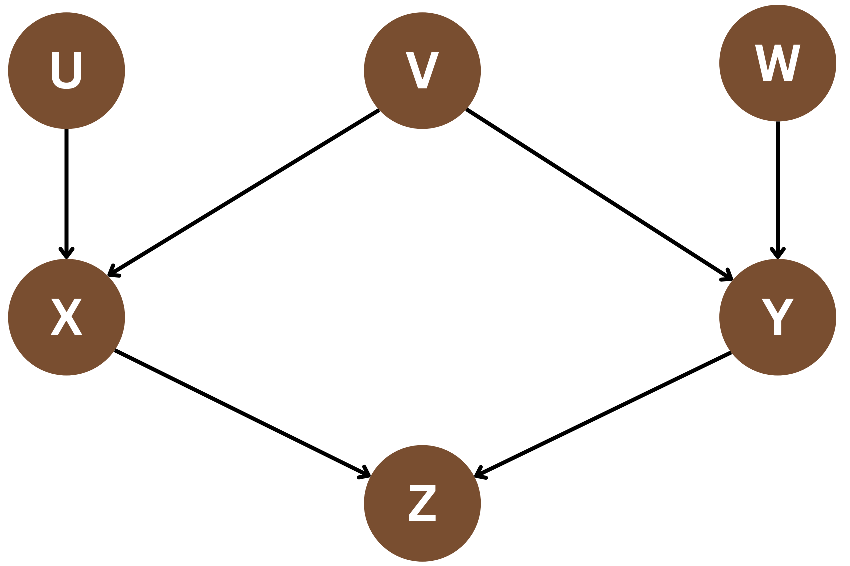
Factores Iniciales

$$P(U); P(V); P(W); P(X \mid U, V); P(Y \mid V, W); P(+z \mid X, Y)$$



$$W, Y, V, X$$

Inferencia: Variable Elimination



Query $P(U \mid +z)$

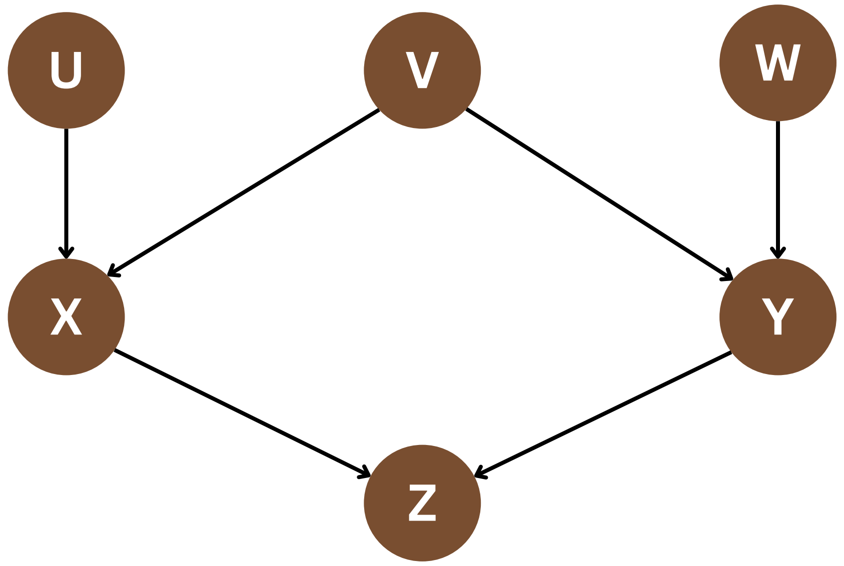
Factores

$$P(U); P(V); \boxed{P(W)}; P(X \mid U, V); \boxed{P(Y \mid V, W)}; P(+z \mid X, Y)$$

$$W, Y, V, X$$

Eliminar W $f_1(Y, V) = \sum_w P(w) P(Y \mid V, w)$

Inferencia: Variable Elimination



$$W, Y, V, X$$

Query $P(U \mid +z)$

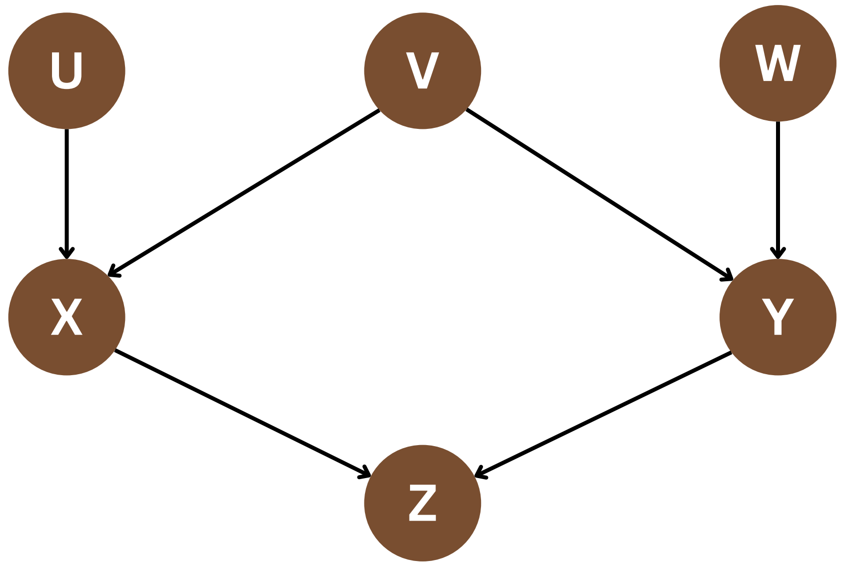
Factores

$$P(U); P(V); P(X \mid U, V); \boxed{P(+z \mid X, Y)}; \boxed{f_1(Y, V)}$$

Eliminar W $$f_1(Y, V) = \sum_w P(w) P(Y \mid V, w)$$

Eliminar Y $$f_2(+z, V, X) = \sum_y f_1(y, V) P(+z \mid X, y)$$

Inferencia: Variable Elimination



W, Y, V, X

Query $P(U \mid +z)$

Factores

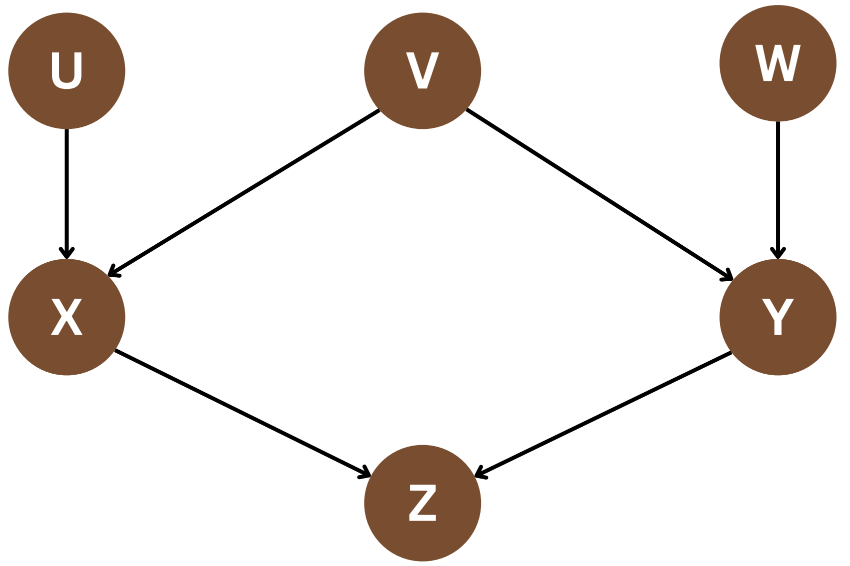
$$P(U); \boxed{P(V)}; \boxed{P(X \mid U, V)}; \boxed{f_2(+z \mid V, X)}$$

Eliminar W $$f_1(Y, V) = \sum_w P(w) P(Y \mid V, w)$$

Eliminar Y $$f_2(+z, V, X) = \sum_y f_1(y, V) P(+z \mid X, y)$$

Eliminar V $$f_3(+z, X, U) = \sum_v P(v) P(X \mid U, v) f_2(+z, v, X)$$

Inferencia: Variable Elimination



$$W, Y, V, X$$

Query $P(U \mid +z)$

Factores

$$\boxed{P(U); f_3(+z, X, U)}$$

Eliminar W $f_1(Y, V) = \sum_w P(w) P(Y \mid V, w)$

Eliminar Y $f_2(+z, V, X) = \sum_y f_1(y, V) P(+z \mid X, y)$

Eliminar V $f_3(+z, X, U) = \sum_v P(v) P(X \mid U, v) f_2(+z, v, X)$

Eliminar X $f_4(+z, U) = \sum_x f_3(+z, x, U)$

MINERÍA DE DATOS

Maximiliano Ojeda

muojeda@uc.cl