Probabilistic Graphical Models

Lecture 5

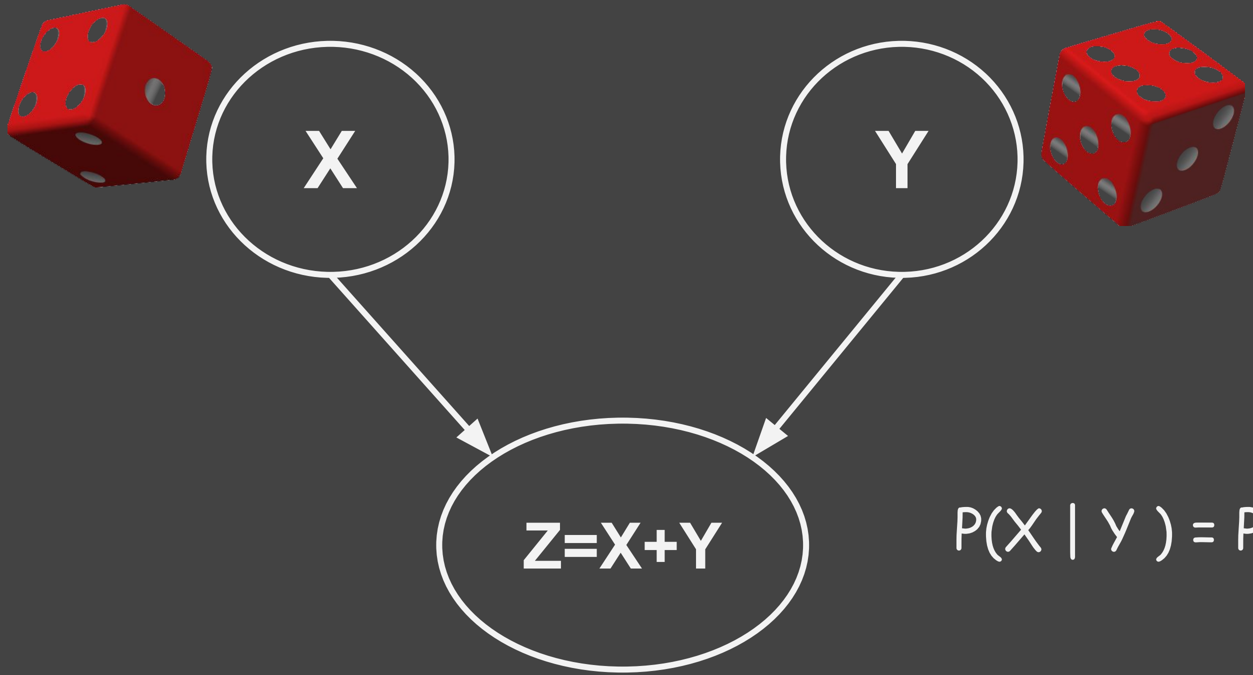
Bayesian Networks

Conditioning can destroy independence

$$P(X \mid Y) = P(X)$$

1928
K. N. Toosi University of Technology

Conditioning can destroy independence

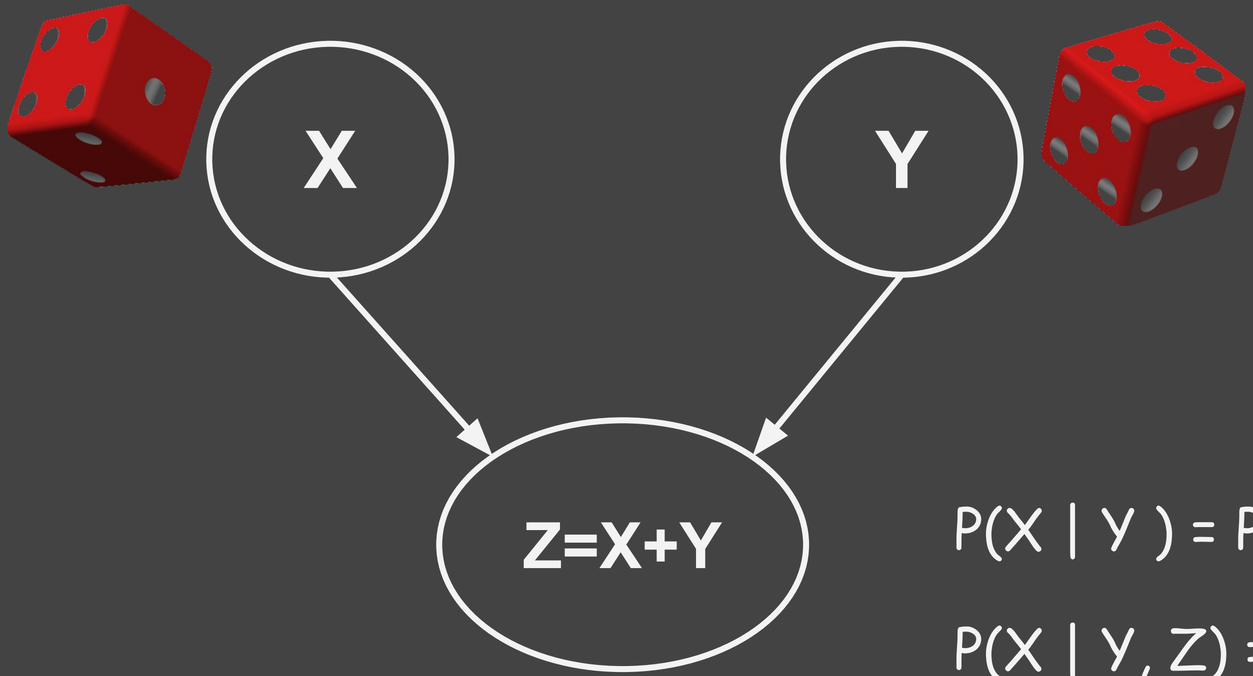


$$P(X \mid Y) = P(X)$$

1928
K. N. Toosi University of Technology

Conditioning can destroy independence

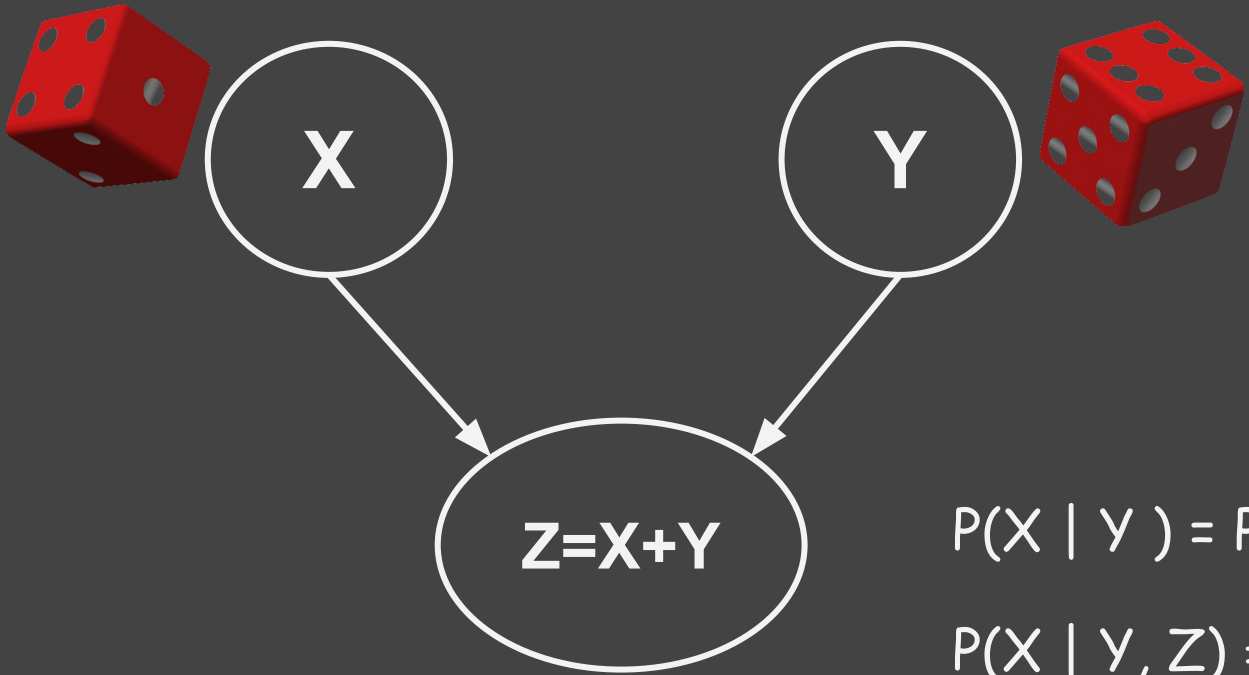
1928
K. N. Toosi University of Technology



$$P(X \mid Y) = P(X)$$

$$P(X \mid Y, Z) = P(X \mid Z)\ ?$$

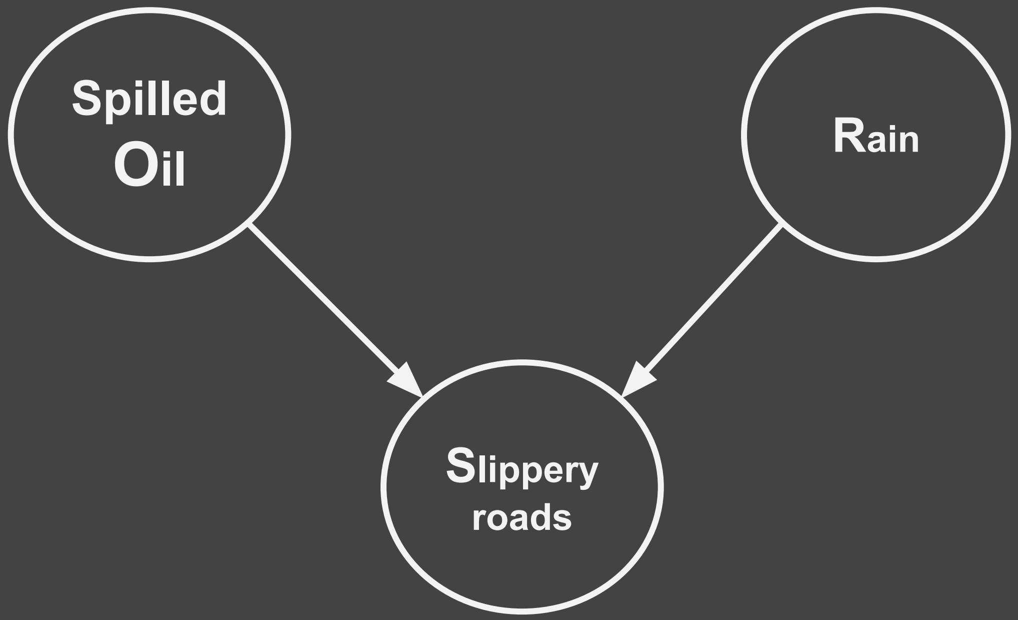
Conditioning can destroy independence



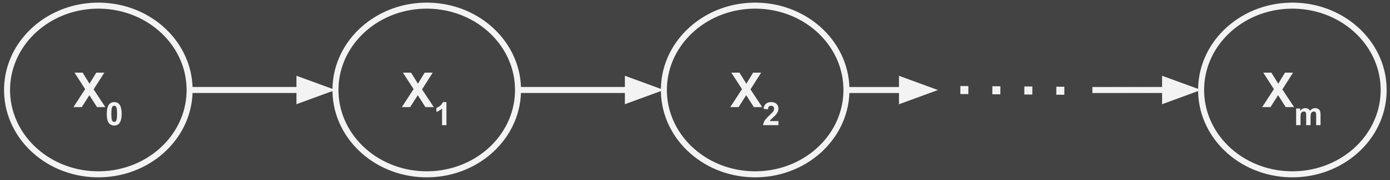
$$P(X \mid Y) = P(X)$$

$$P(X \mid Y, Z) \neq P(X \mid Z)$$

Conditioning can destroy independence



Remember: faulty push-button

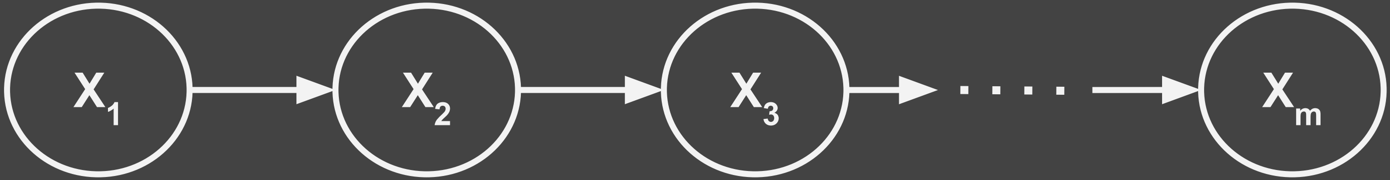


- x_n device on or off (1/0) after n times pressing the button
- button works with probability p_t if the device is on, and with probability q_t if the device is off.

$p(x_1, x_2, \ldots, x_m)$ $\quad x_i \in 0, 1$ (2^m-1 parameters)

$p(x_1, x_2, \ldots, x_m) = p(x_1) \cdots p(x_m)$ (m parameters)

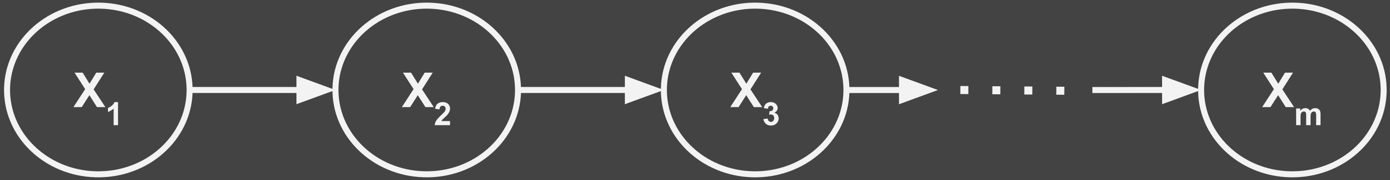
Remember: faulty push-button



Observation: $p(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_2, x_1) = p(x_t \mid x_{t-1})$

$$p(x_1, x_2, \ldots, x_m) = p(x_m \mid x_1, x_2, \ldots, x_{m-1})\, p(x_1, x_2, \ldots, x_{m-1})$$

Remember: faulty push-button



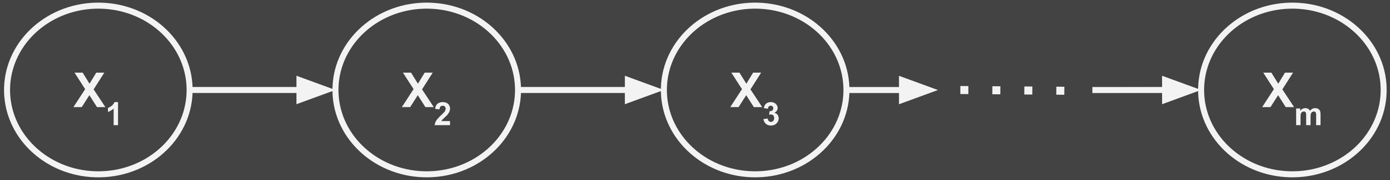
Observation: $p(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_2, x_1) = p(x_t \mid x_{t-1})$

$$p(x_1, x_2, \ldots, x_m) = p(x_m \mid x_1, x_2, \ldots, x_{m-1})\, p(x_1, x_2, \ldots, x_{m-1})$$

$$= p(x_m \mid x_{m-1})\, p(x_1, x_2, \ldots, x_{m-1})$$

Remember: faulty push-button

K. N. Toosi
University of Technology



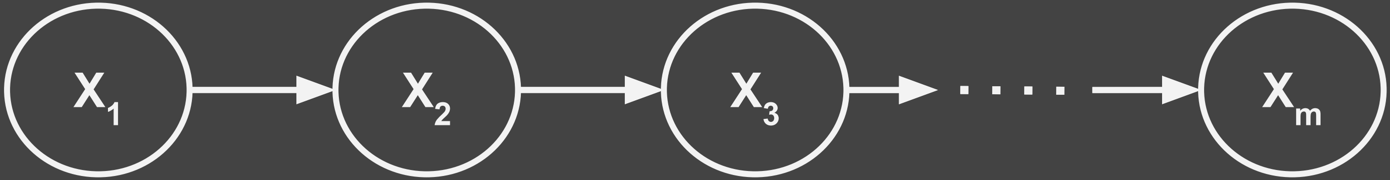
Observation: $p(x_t \mid x_{t-1}, x_{t-2}, \dots, x_2, x_1) = p(x_t \mid x_{t-1})$

$$p(x_1, x_2, \dots, x_m) = p(x_m \mid x_1, x_2, \dots, x_{m-1})\, p(x_1, x_2, \dots, x_{m-1})$$

$$= p(x_m \mid x_{m-1})\, p(x_1, x_2, \dots, x_{m-1})$$

$$= p(x_m \mid x_{m-1})\, p(x_{m-1} \mid x_1, x_2, \dots, x_{m-2})\, p(x_1, x_2, \dots, x_{m-2})$$

Remember: faulty push-button



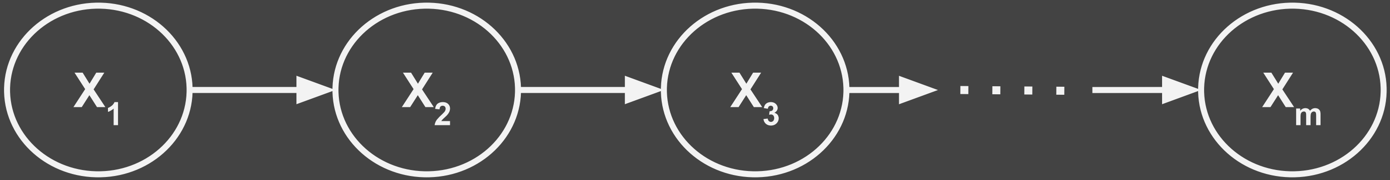
Observation: $p(x_t \mid x_{t-1}, x_{t-2}, \dots, x_2, x_1) = p(x_t \mid x_{t-1})$

$$p(x_1, x_2, \dots, x_m) = \dots = p(x_m \mid x_{m-1})\, p(x_1, x_2, \dots, x_{m-1})$$

$$= p(x_m \mid x_{m-1})\, p(x_{m-1} \mid x_1, x_2, \dots, x_{m-2})\, p(x_1, x_2, \dots, x_{m-2})$$

$$= p(x_m \mid x_{m-1})\, p(x_{m-1} \mid x_{m-2})\, p(x_1, x_2, \dots, x_{m-2})$$

Remember: faulty push-button

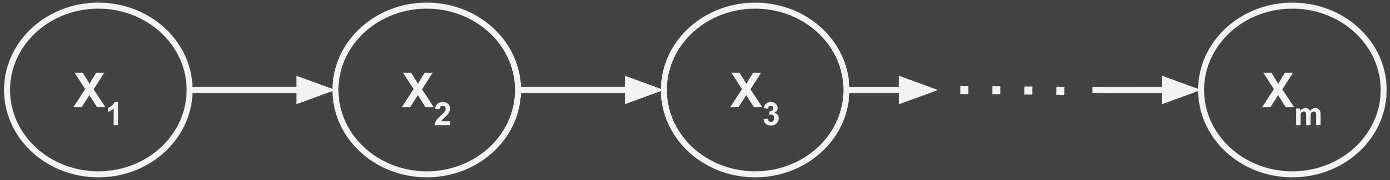


Observation: $p(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_2, x_1) = p(x_t \mid x_{t-1})$

$$p(x_1, x_2, \ldots, x_m) = \ldots$$

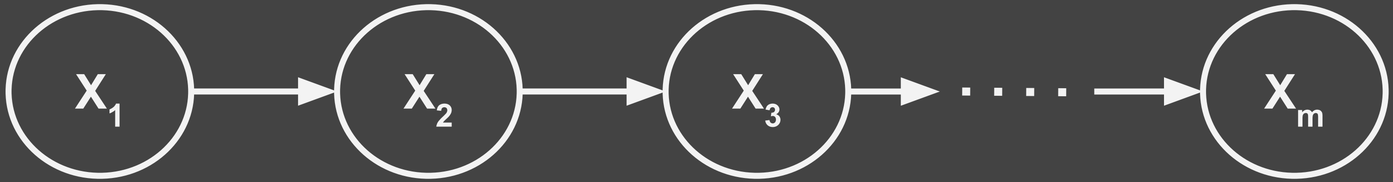
$$= p(x_1) \; p(x_2 | x_1) \; p(x_3 | x_2) \; \ldots \; p(x_{m-1} | x_{m-2}) \; p(x_m | x_{m-1})$$

Remember: faulty push-button



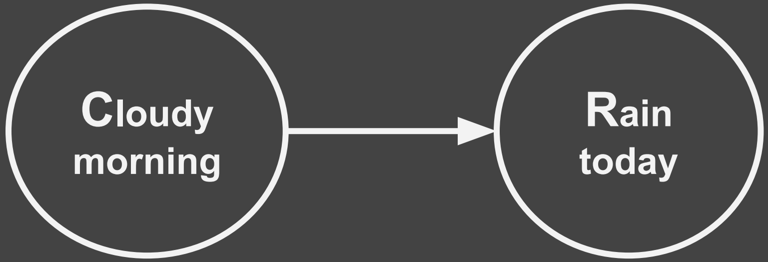
- fully dependent: 2^m-1 free parameters (about 10^{30} for n=100)
- fully independent: m free parameters (100 for n=100)
  - $p(x_1, x_2, \ldots, x_m) = p(x_1)\ p(x_2)\ \ldots\ p(x_m)$
- conditionally independent: 2m-1 free parameters (199 for n=100)
  - $p(x_1, x_2, \ldots, x_m) = p(x_1)\ \ p(x_2\,|x_1)\ \ \ldots\ p(x_m\,|x_{m-1})$

Remember: faulty push-button



Given X_{t-1}, X_t is independent of $X_1, \ldots, X_{t-2}$

Directed graphs



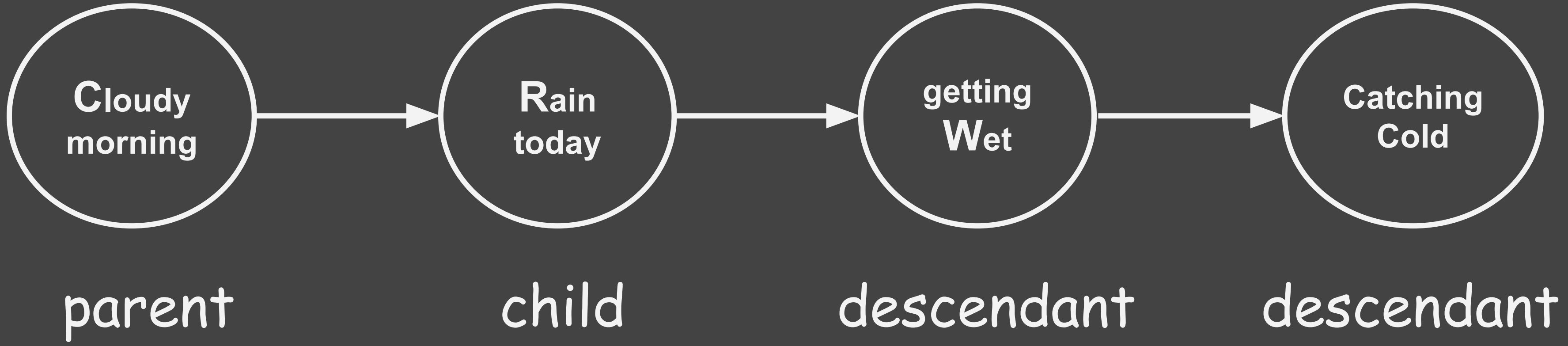
parent child

Directed graphs

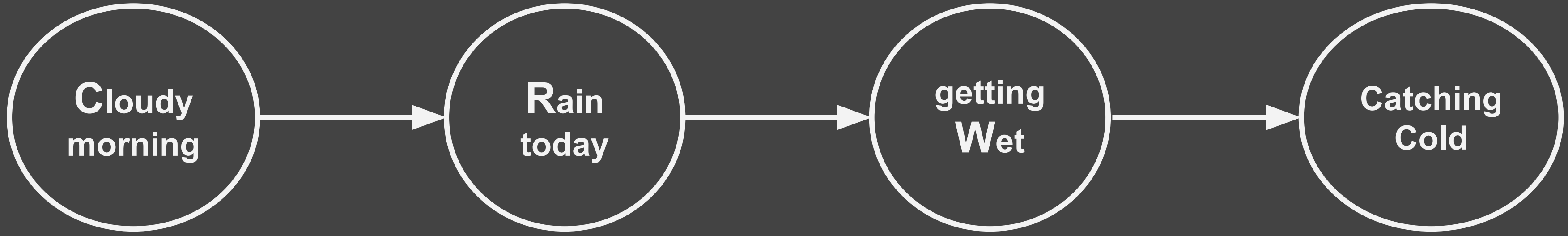
Directed graphs

parent child descendant

Directed graphs



Directed graphs



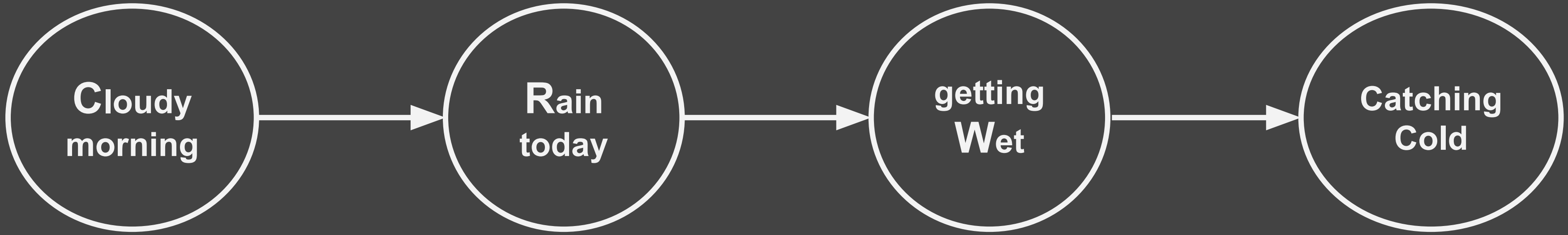
parent

child
descendant

descendant

descendant

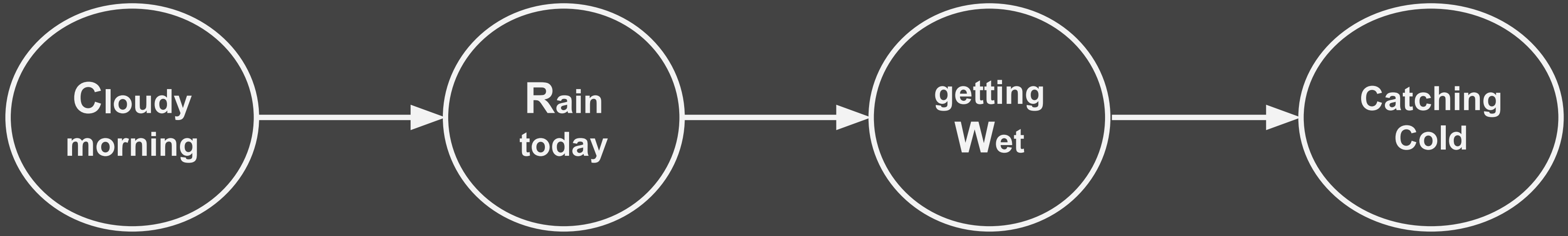
Directed graphs



ancestor ancestor parent child

Directed Graphs & Conditional Independence

K. N. Toosi
University of Technology



ancestor ancestor parent child

given (immediate) parent, child is independent of all ancestors

Directed Graphs & Conditional Independence

$$P(C,R,W) = P(W \mid C, R)\, P(C,R)$$

Directed Graphs & Conditional Independence

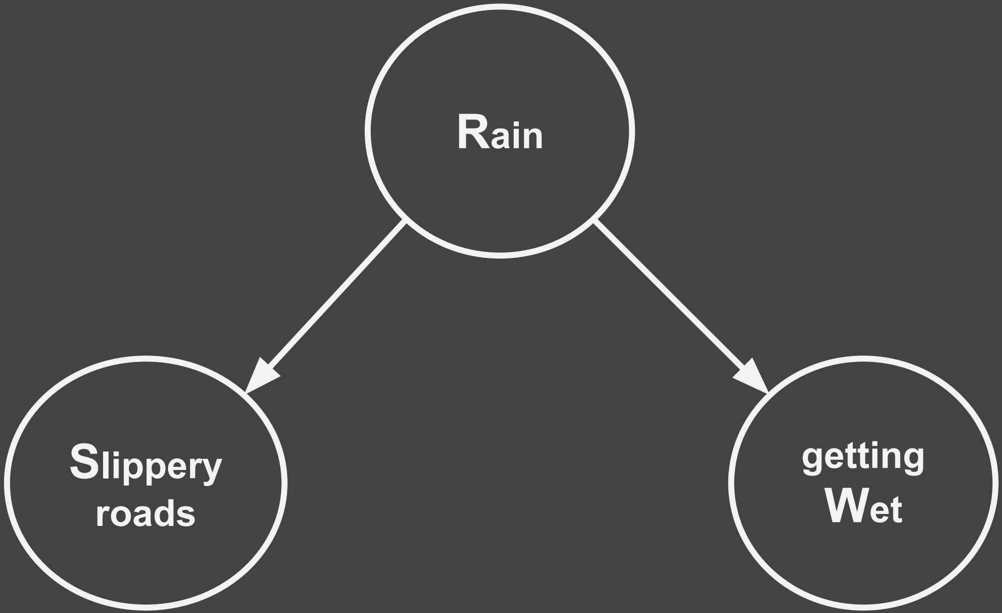
$$P(C,R,W) = P(W \mid C, R)\, P(C,R) = P(W \mid R)\, P(C,R)$$

Directed Graphs & Conditional Independence

$$P(C,R,W) = P(W \mid C, R)\, P(C,R) = P(W \mid R)\, P(C,R)$$

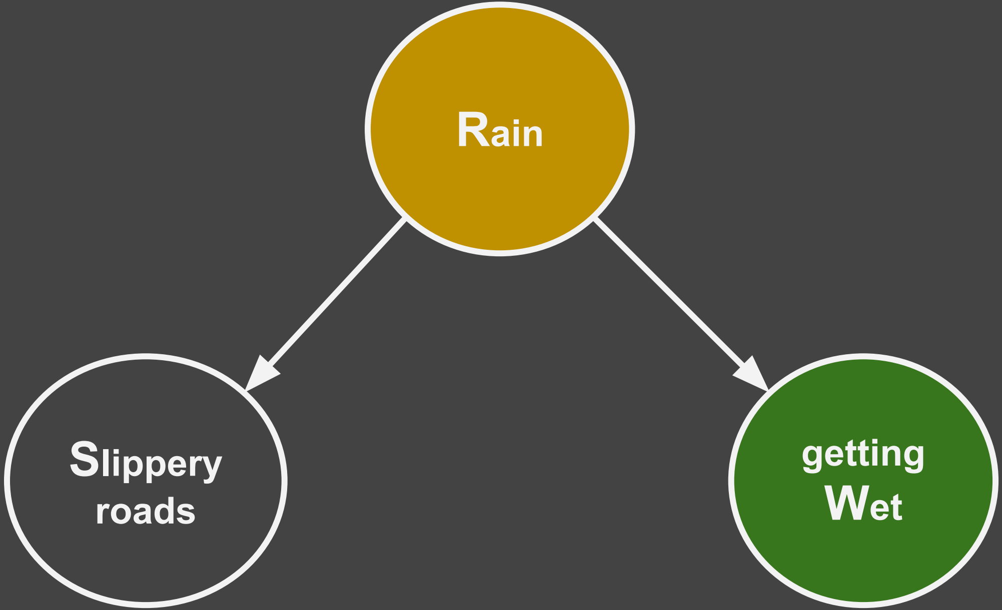
$$= P(W \mid R)\, P(R \mid C)\, P(C)$$

Directed Graphs & Conditional Independence



Directed Graphs & Conditional Independence

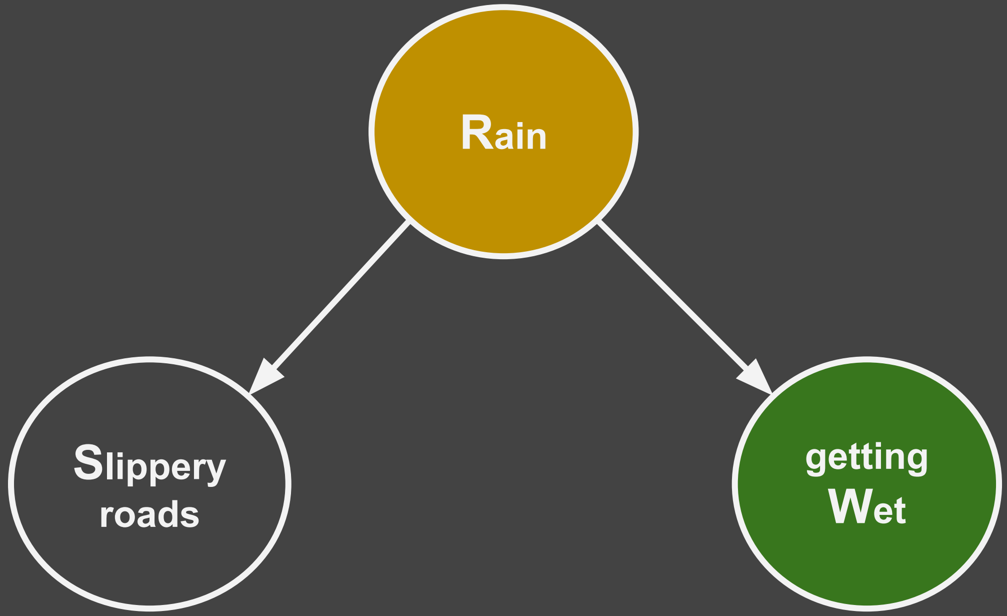
K. N. Toosi
University of Technology



given (immediate) parent, child is independent of ?

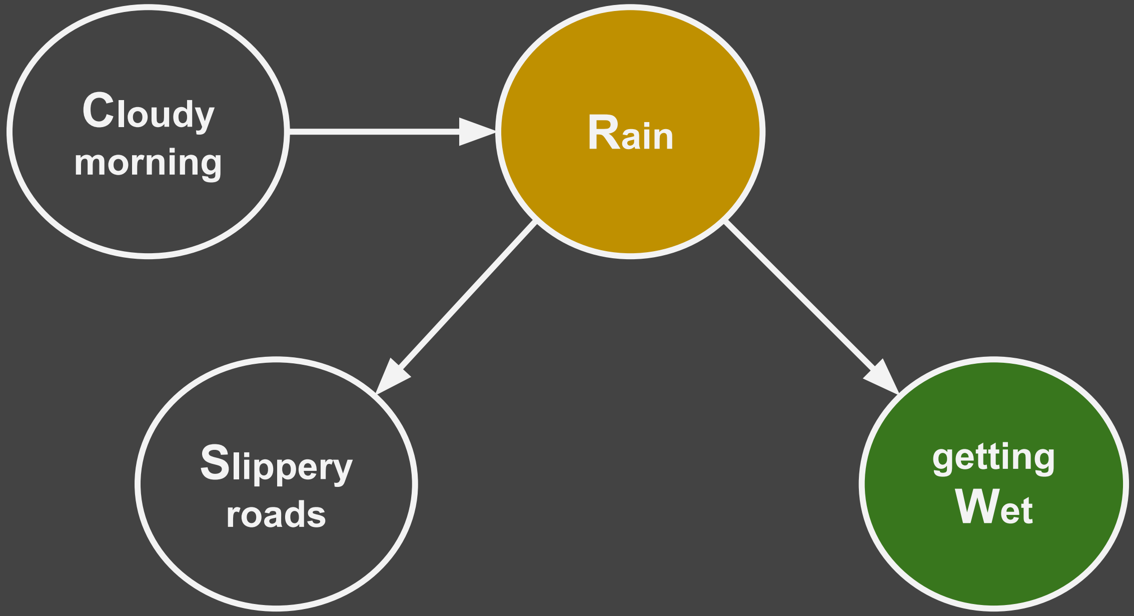
Directed Graphs & Conditional Independence

K. N. Toosi
University of Technology



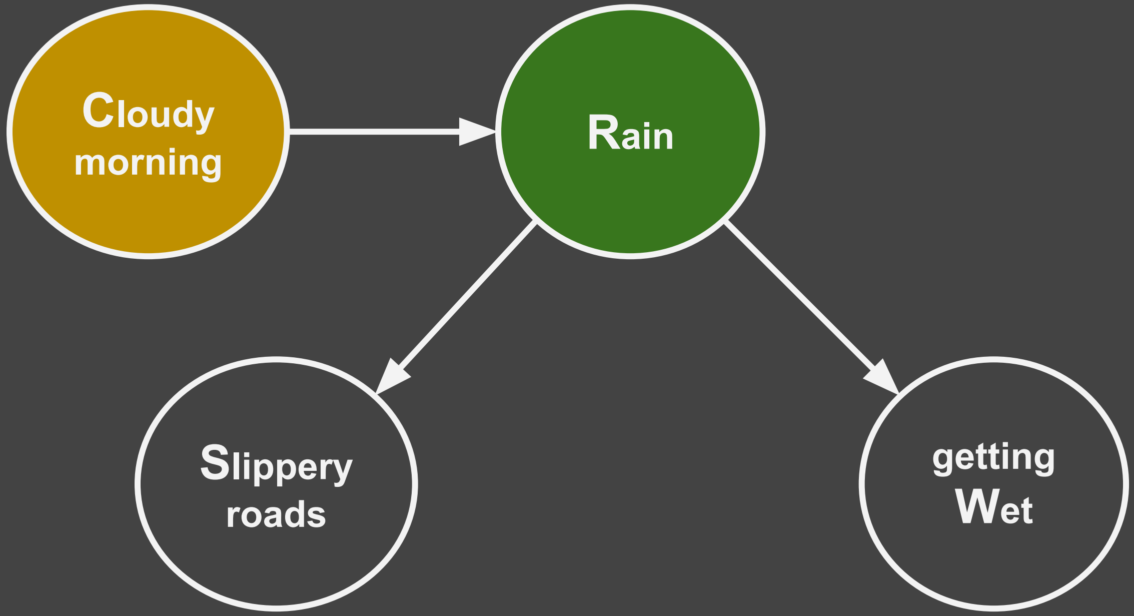
given (immediate) parent, child is independent of all other nodes?

Directed Graphs & Conditional Independence



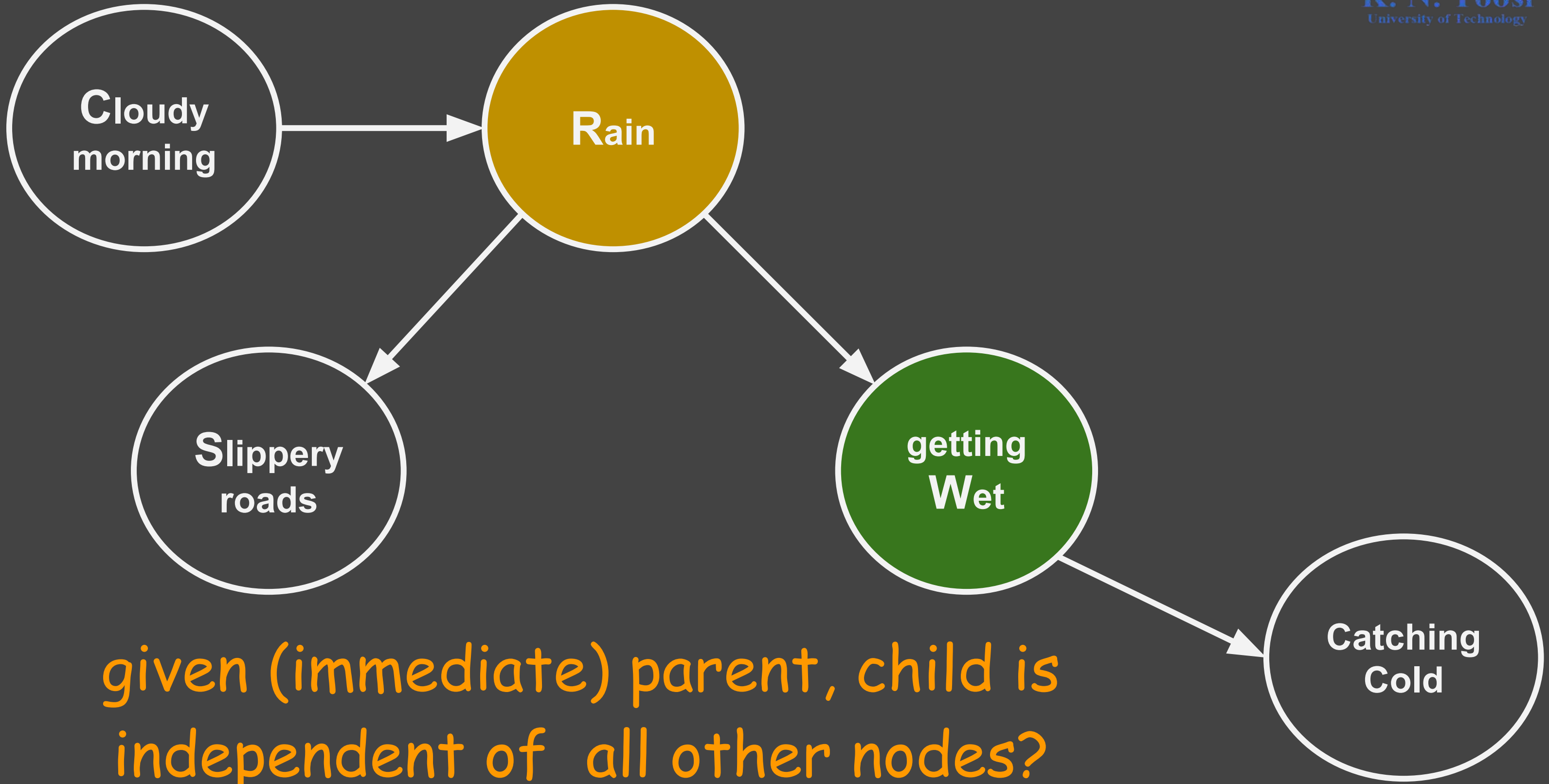
given (immediate) parent, child is independent of all other nodes?

Directed Graphs & Conditional Independence



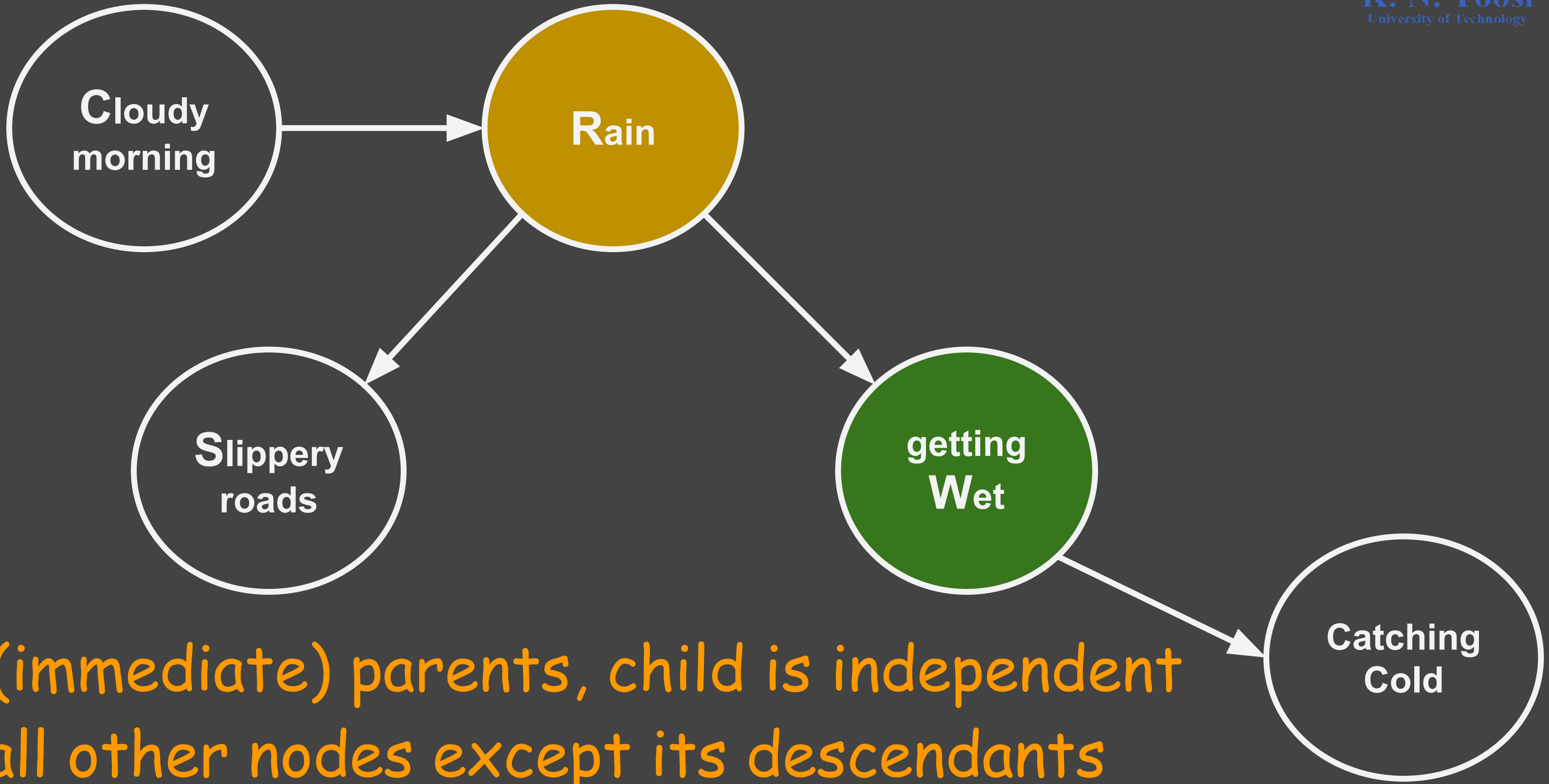
given (immediate) parent, child is independent of all other nodes?

Directed Graphs & Conditional Independence



given (immediate) parent, child is independent of all other nodes?

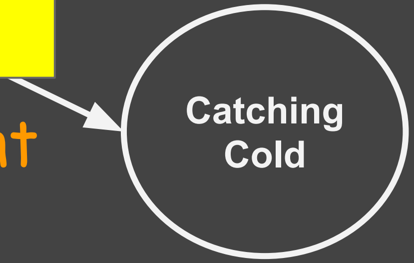
Directed Graphs & Conditional Independence



given (immediate) parents, child is independent of all other nodes except its descendants

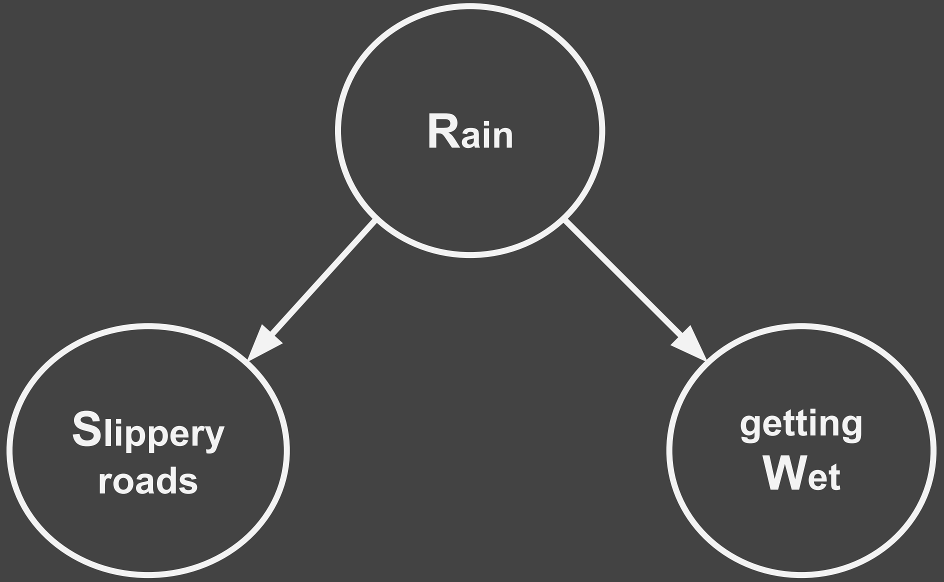
Directed Graphs & Conditional Independence

Markov Property



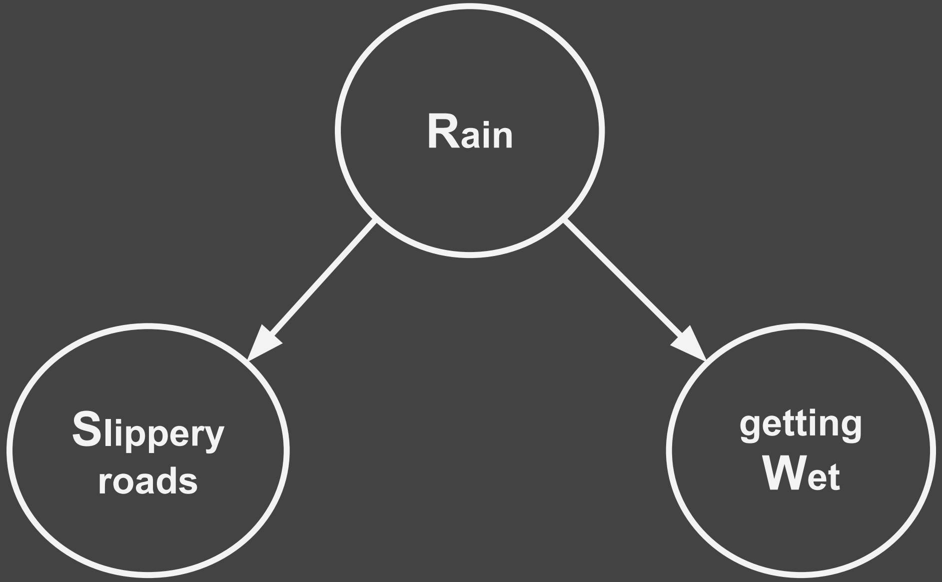
given (immediate) parents, child is independent of all other nodes except its descendants

Directed Graphs & Conditional Independence



$$P(R,S,W) = P(W \mid R,S)\, P(R,S)$$

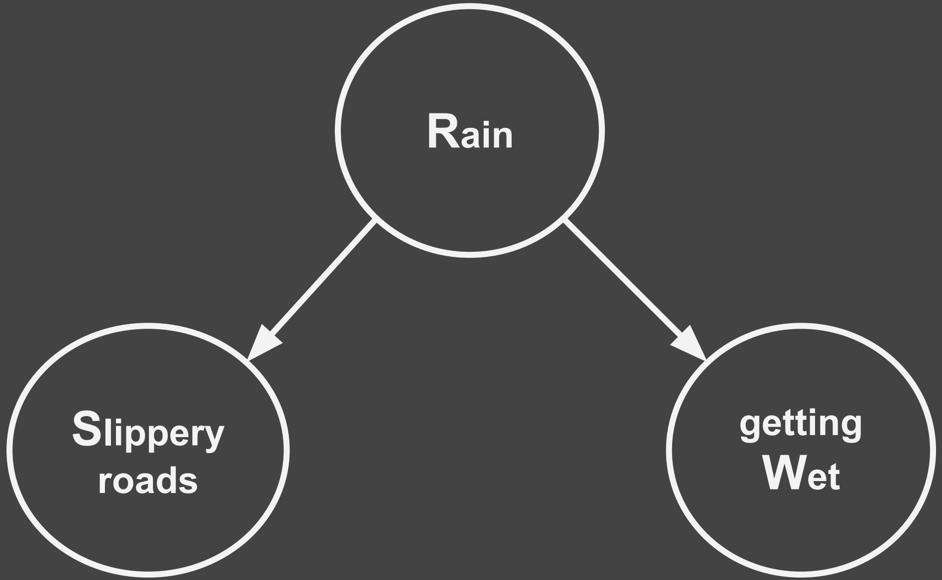
Directed Graphs & Conditional Independence



$$P(R,S,W) = P(W \mid R,S)\, P(R,S) = P(W \mid R)\, P(R,S)$$

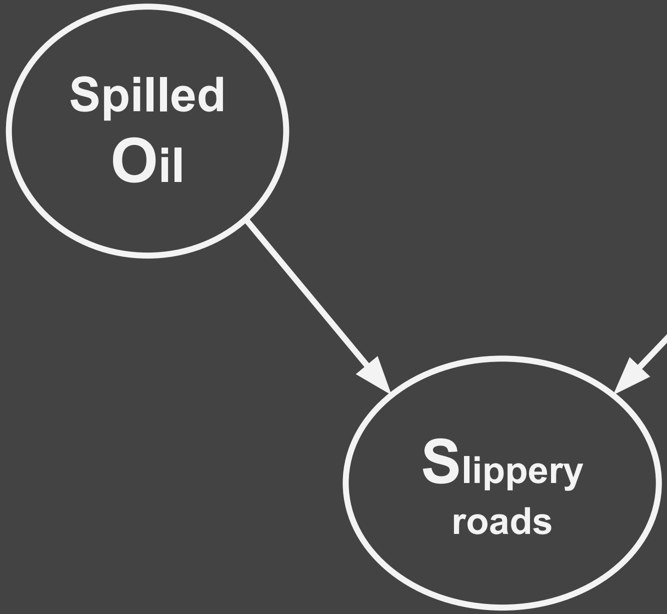
Directed Graphs & Conditional Independence

K. N. Toosi
University of Technology



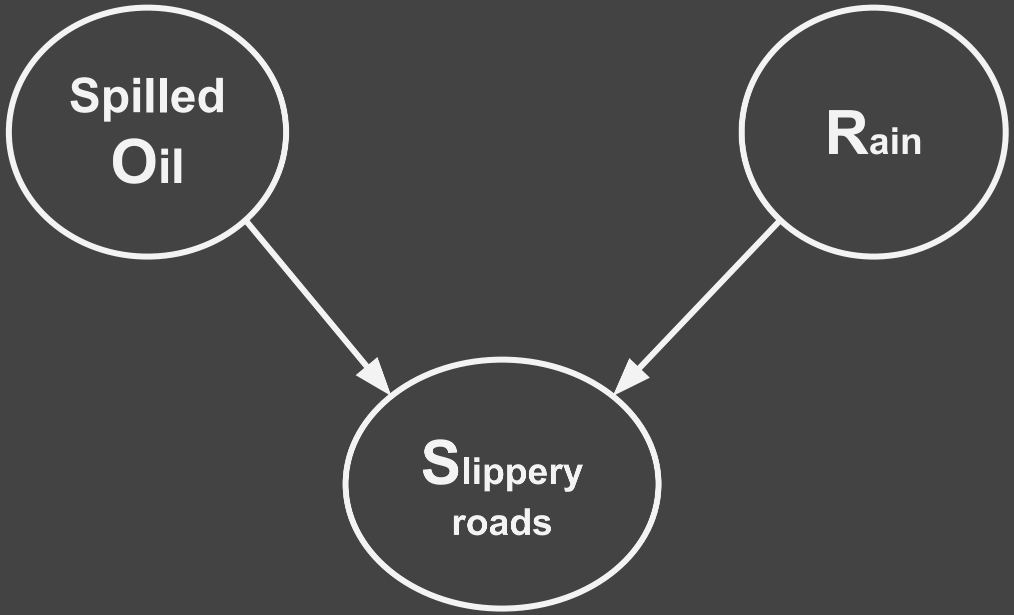
$$P(R,S,W) = P(W \mid R,S)\, P(R,S) = P(W \mid R)\, P(R,S)$$
$$= P(W \mid R)\, P(S|R)\, P(R)$$

Directed Graphs & Conditional Independence



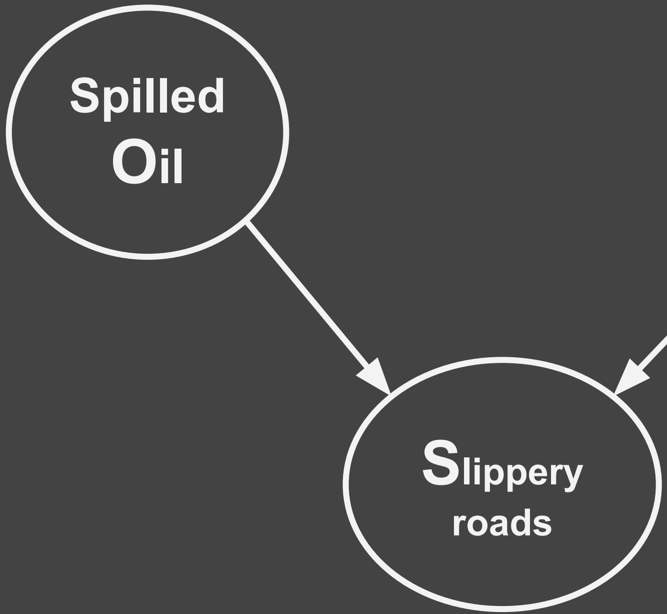
$P(O,R,S) =$

Directed Graphs & Conditional Independence



$$P(O,R,S) = P(S \mid O, R)\, P(O, R)$$

Directed Graphs & Conditional Independence

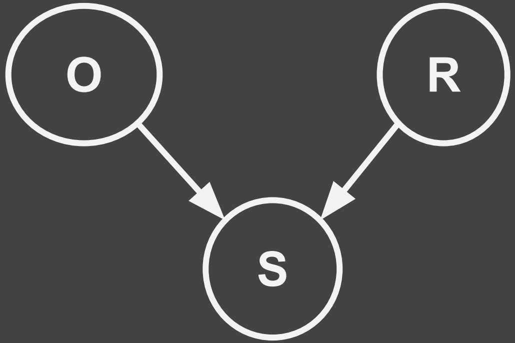


$$P(O,R,S) = P(S \mid O, R)\, P(O, R)$$
$$= P(S \mid O, R)\, P(O)\, P(R)$$

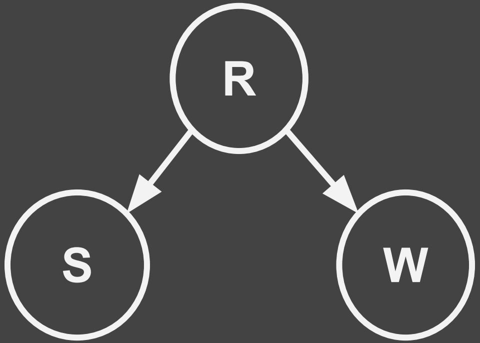
Directed Graphs & Conditional Independence

K. N. Toosi
University of Technology

$$P(C,R,W) = P(W \mid R)\, P(R \mid C)\, P(C)$$

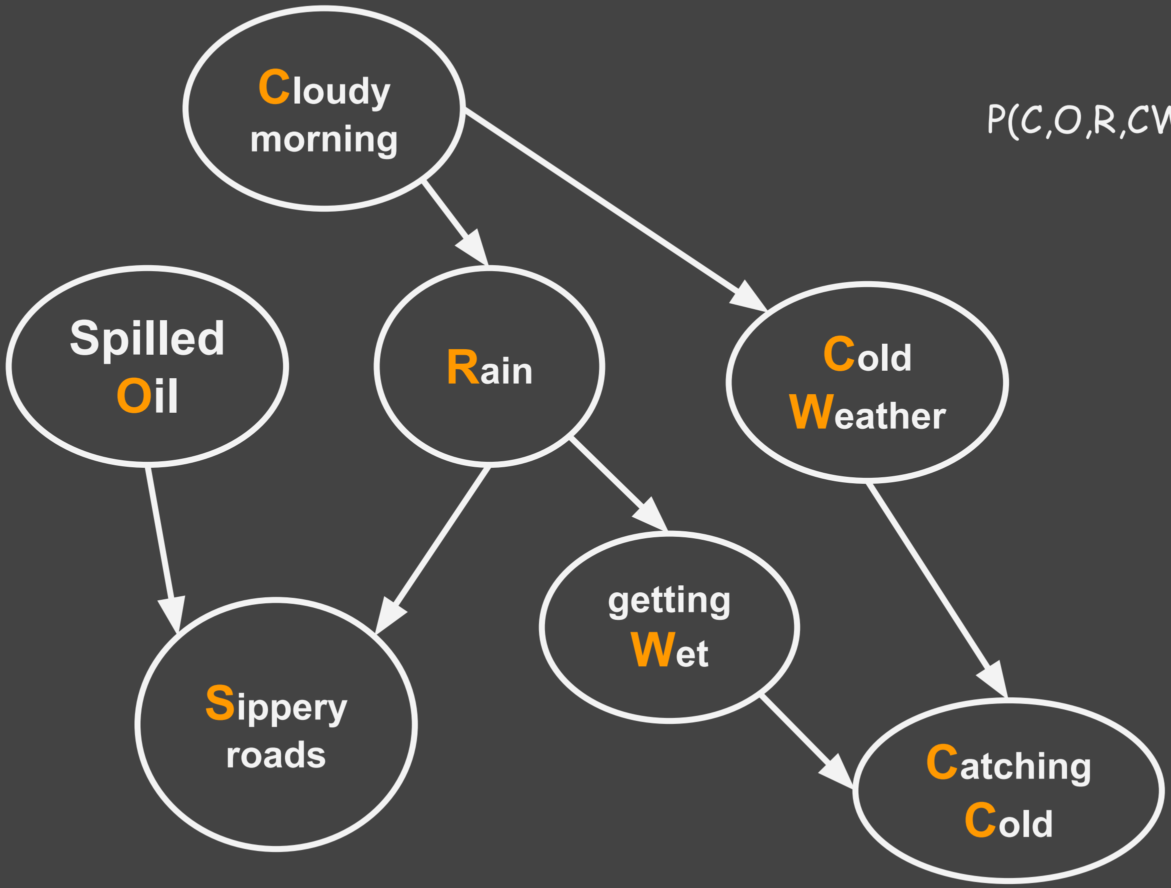


$$P(O,R,S) = P(S \mid O, R)\, P(O)\, P(R)$$



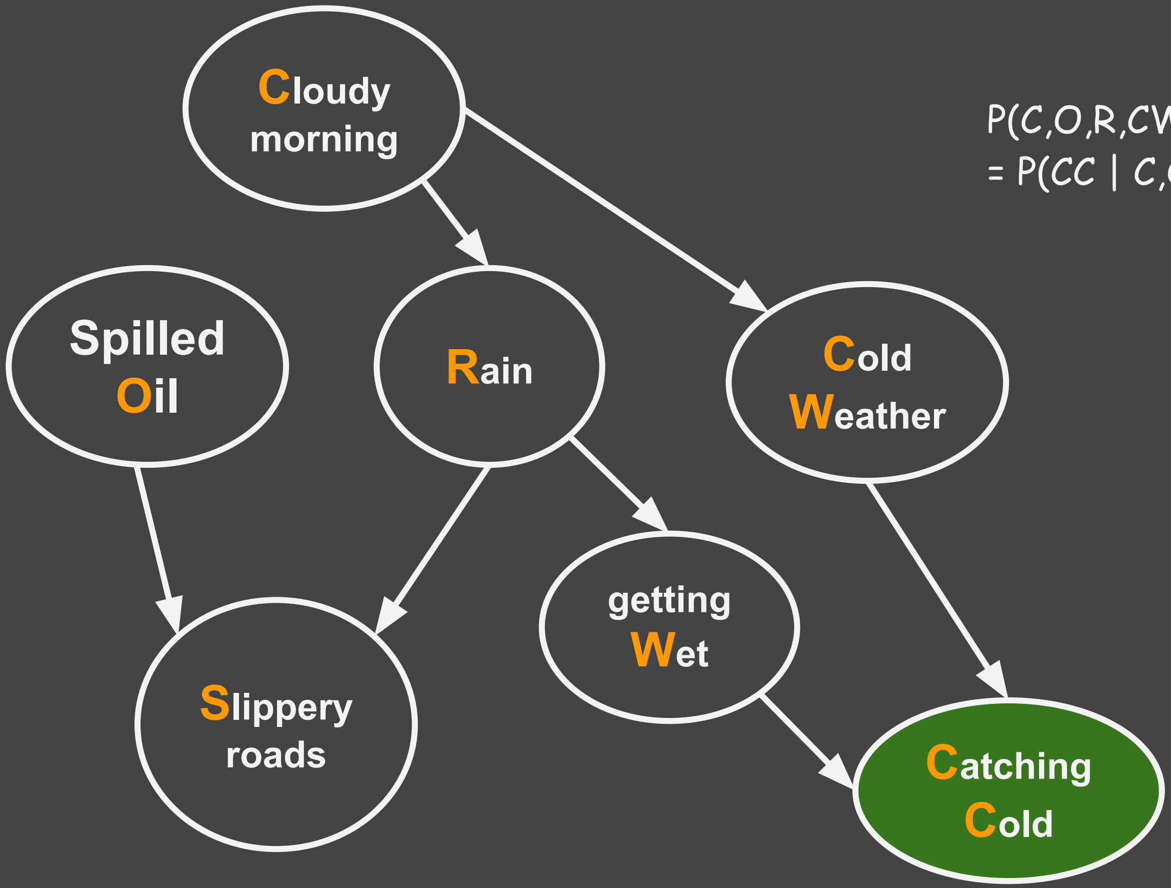
$$P(R,S,W) = P(W \mid R)\, P(S \mid R)\, P(S)$$

Directed Graphs & Conditional Independence



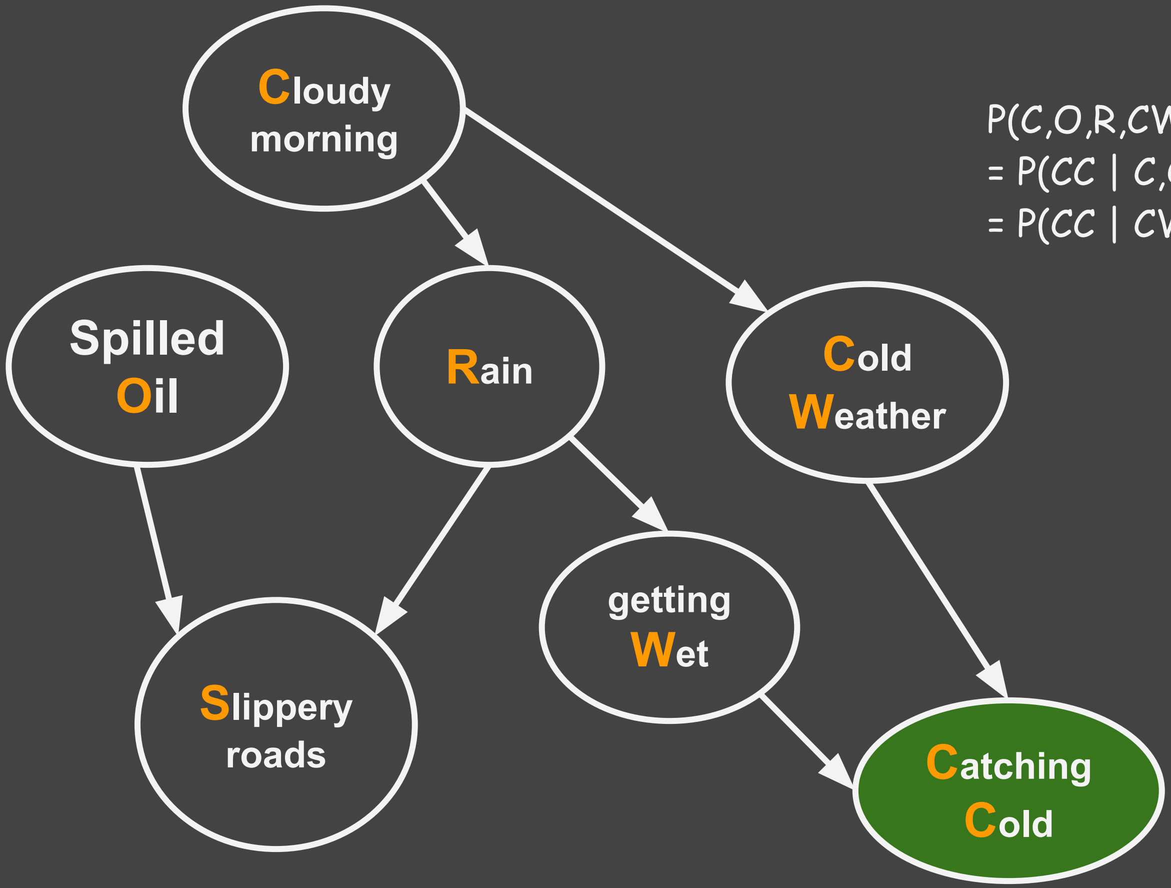
P(C,O,R,CW,S,W,CC) = ?

Directed Graphs & Conditional Independence



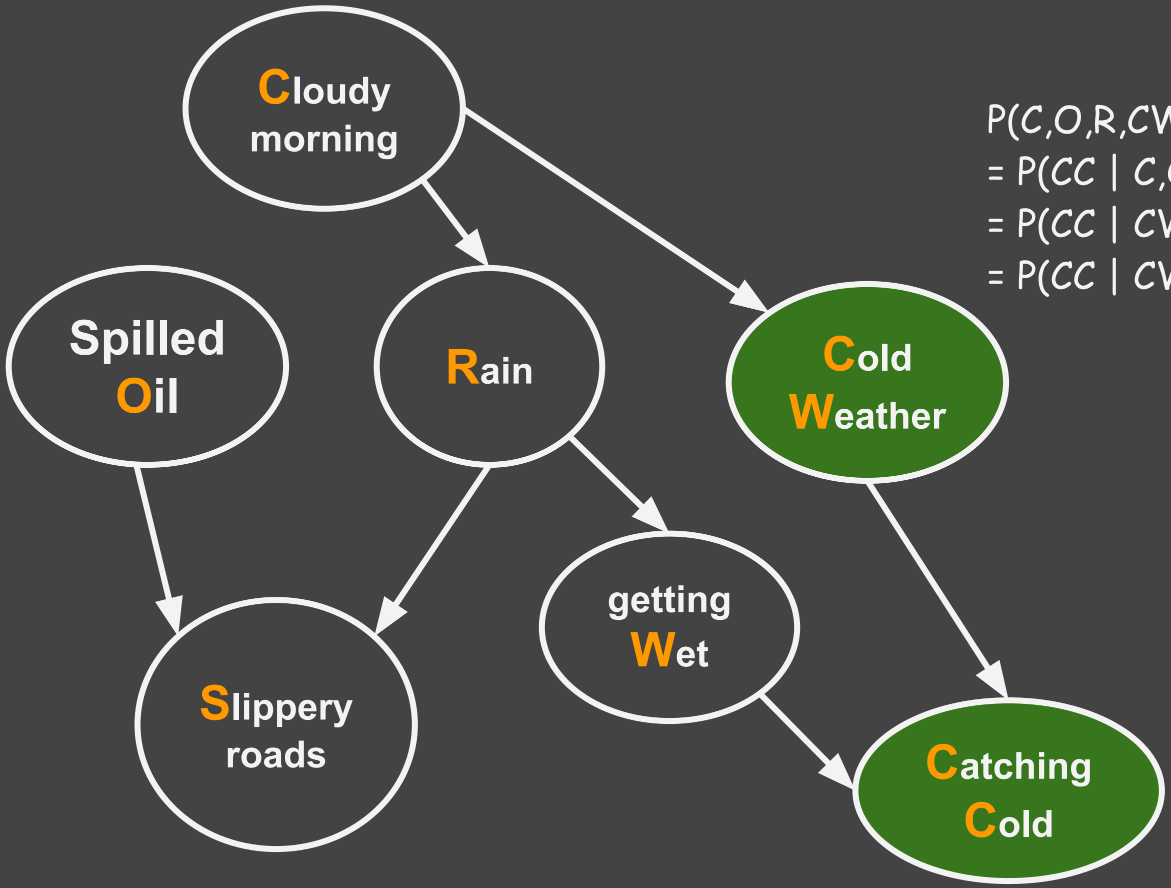
$P(C,O,R,CW,S,W,CC)$
$= P(CC \mid C,O,R,CW,S,W)\, P(C,O,R,CW,S,W)$

Directed Graphs & Conditional Independence



$$P(C,O,R,CW,S,W,CC)$$
$$= P(CC \mid C,O,R,CW,S,W)\, P(C,O,R,CW,S,W)$$
$$= P(CC \mid CW,W)\, P(C,O,R,CW,S,W)$$

Directed Graphs & Conditional Independence



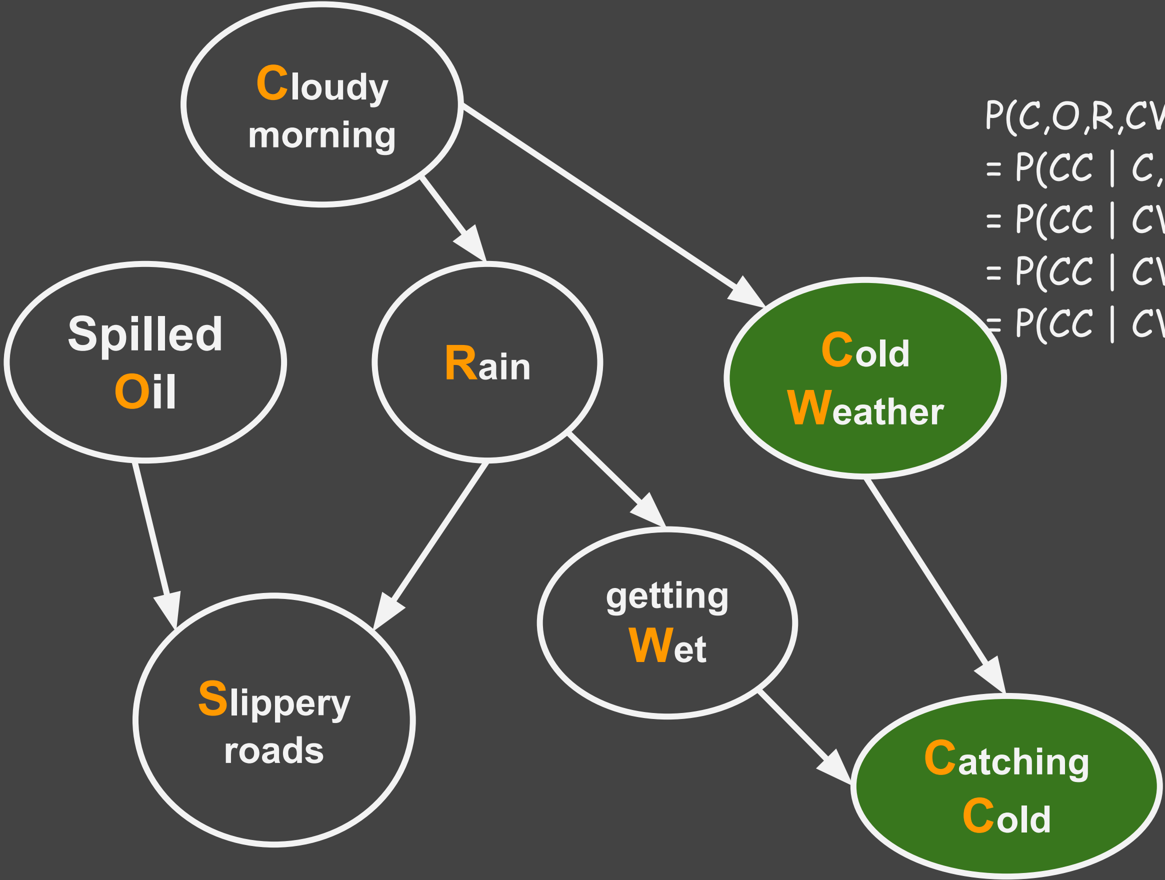
P(C,O,R,CW,S,W,CC)

= P(CC | C,O,R,CW,S,W) P(C,O,R,CW,S,W)

= P(CC | CW,W) P(C,O,R,CW,S,W)

= P(CC | CW,W) P(CW|C,O,R,S,W) P(C,O,R,S,W)

Directed Graphs & Conditional Independence

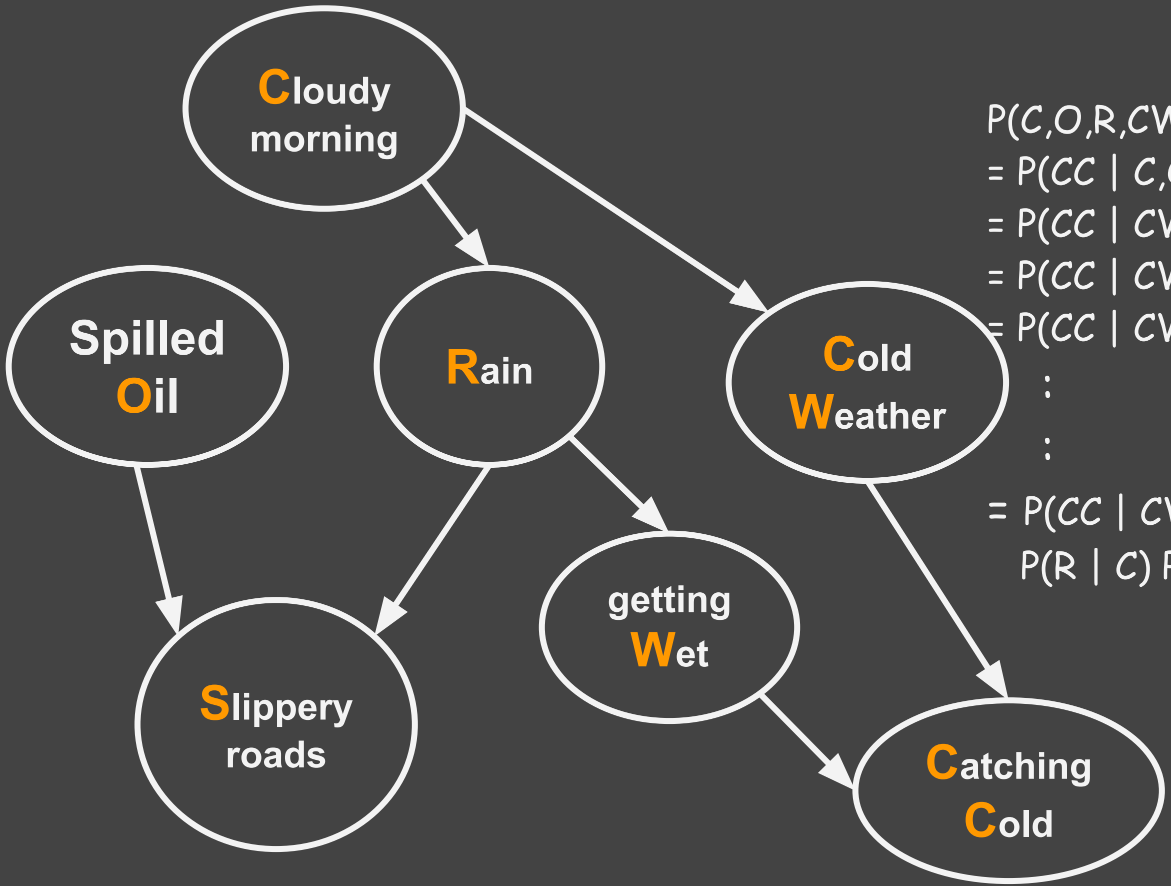


$$
\begin{aligned}
&P(C,O,R,CW,S,W,CC) \\
&= P(CC \mid C,O,R,CW,S,W)\, P(C,O,R,CW,S,W) \\
&= P(CC \mid CW,W)\, P(C,O,R,CW,S,W) \\
&= P(CC \mid CW,W)\, P(CW|C,O,R,S,W)\, P(C,O,R,S,W) \\
&= P(CC \mid CW,W)\, P(CW|C)\, P(C,O,R,S,W)
\end{aligned}
$$

K. N. Toosi
University of Technology

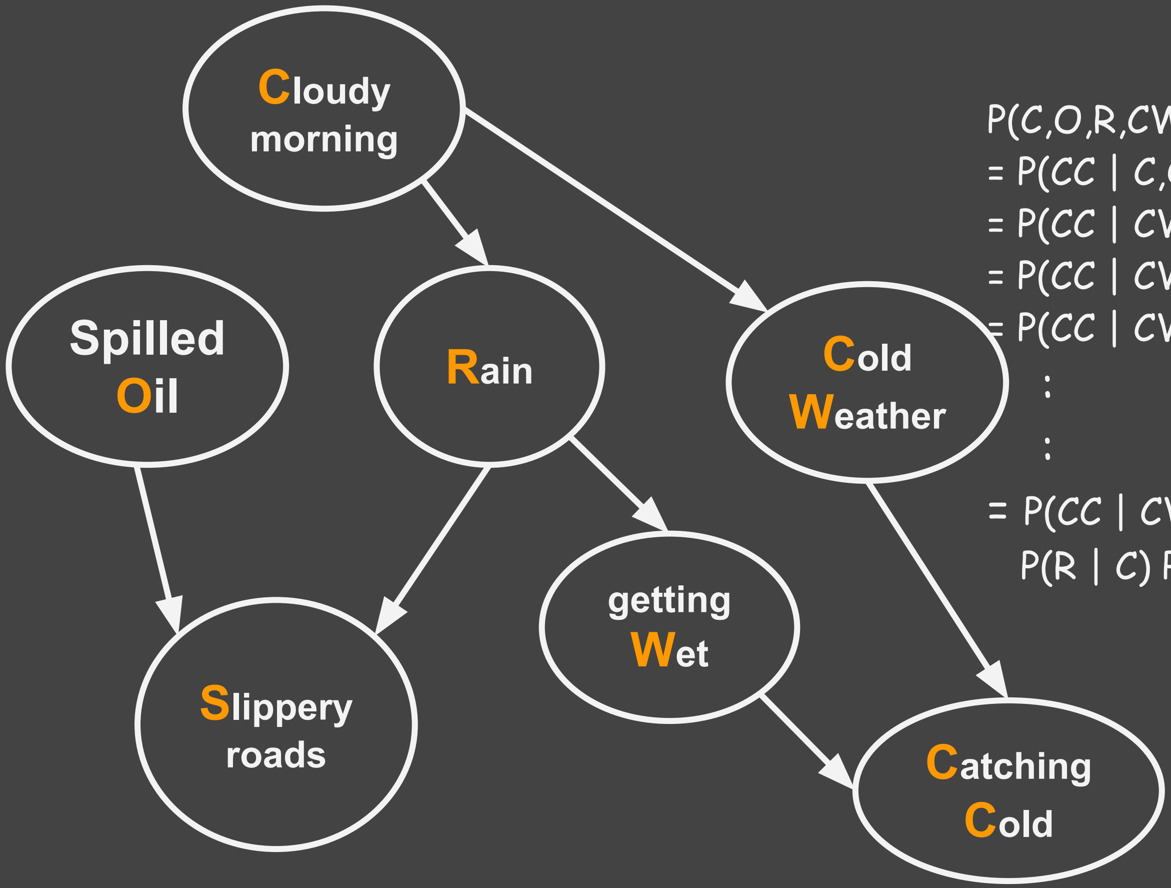
Directed Graphs & Conditional Independence

K. N. Toosi
University of Technology



P(C,O,R,CW,S,W,CC)
= P(CC | C,O,R,CW,S,W) P(C,O,R,CW,S,W)
= P(CC | CW,W) P(C,O,R,CW,S,W)
= P(CC | CW,W) P(CW|C,O,R,S,W) P(C,O,R,S,W)
= P(CC | CW,W) P(CW|C) P(C,O,R,S,W)
⋮
⋮
= P(CC | CW,W) P(CW|C) P(W | R) P(S | O, R) P(R | C) P(O) P(C)

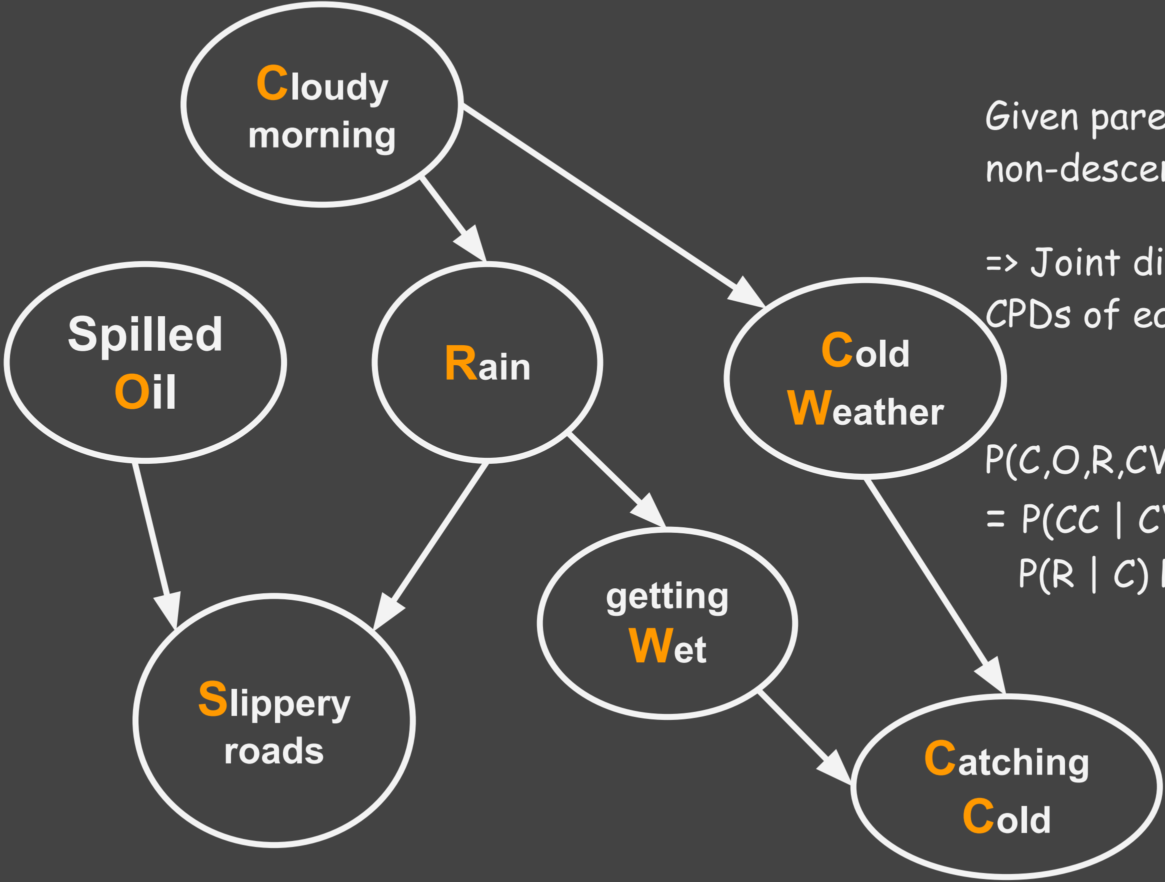
Directed Graphs & Conditional Independence



$$P(C,O,R,CW,S,W,CC)$$
$$= P(CC \mid C,O,R,CW,S,W)\, P(C,O,R,CW,S,W)$$
$$= P(CC \mid CW,W)\, P(C,O,R,CW,S,W)$$
$$= P(CC \mid CW,W)\, P(CW \mid C,O,R,S,W)\, P(C,O,R,S,W)$$
$$= P(CC \mid CW,W)\, P(CW \mid C)\, P(C,O,R,S,W)$$
$$\vdots$$
$$\vdots$$
$$= P(CC \mid CW,W)\, P(CW \mid C)\, P(W \mid R)\, P(S \mid O, R)\, P(R \mid C)\, P(O)\, P(C)$$

Conditional Probability Distributions (CPDs)

Conditional Independence <-> Factorization



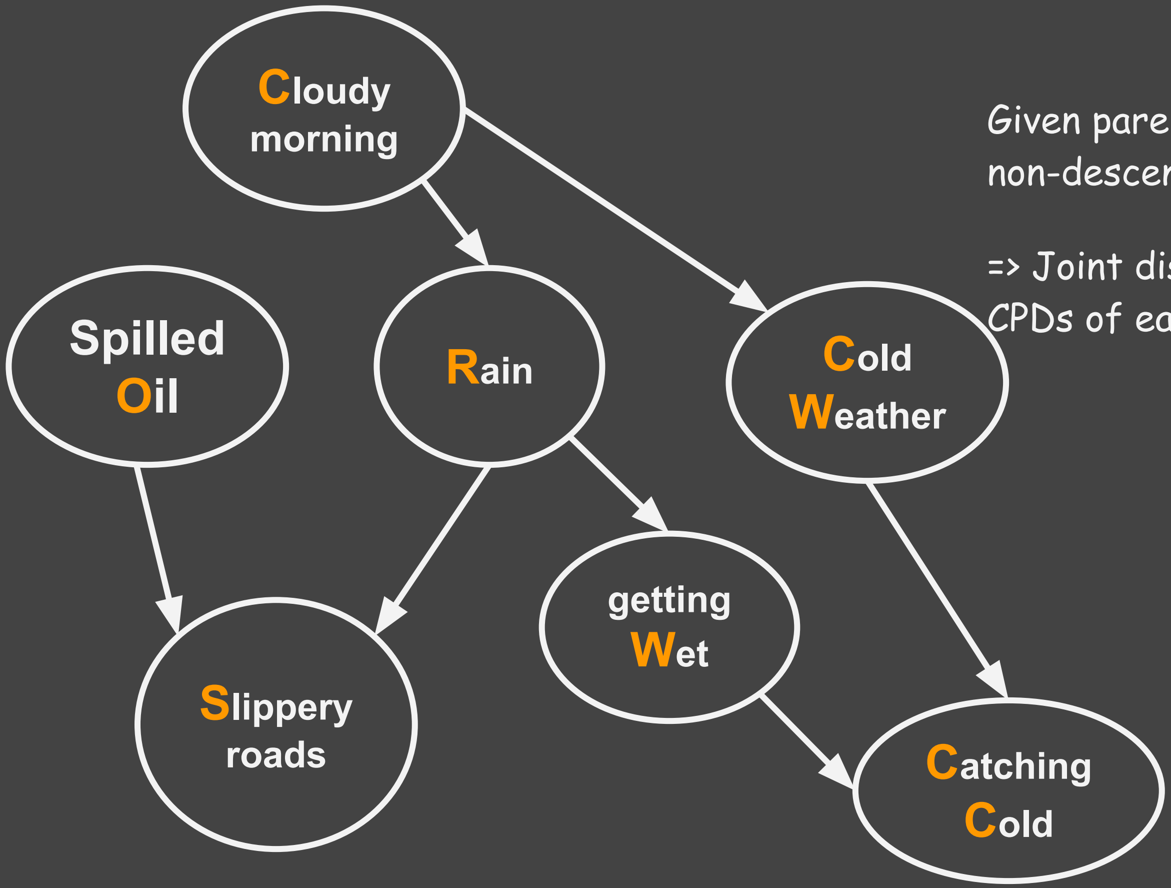
Given parents, nodes are independent of their non-descendants

=> Joint distribution can be factorized into the CPDs of each node given their parents.

P(C,O,R,CW,Sl,W,CC)
= P(CC | CW,W) P(CW|C) P(W | R) P(S | O, R) P(R | C) P(O) P(C)

Conditional Probability Distributions (CPDs)

Conditional Independence <-> Factorization

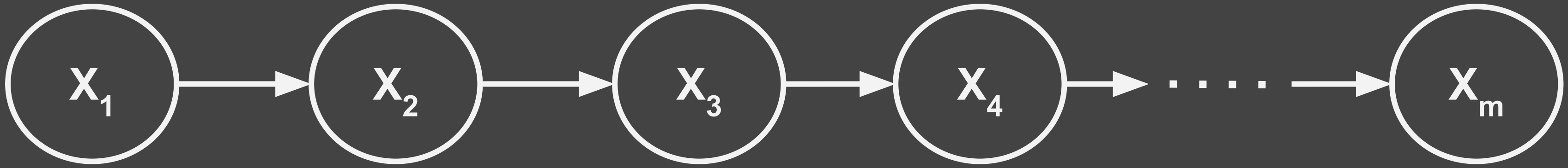


Given parents, nodes are independent of their non-descendants

=> Joint distribution can be factorized into the CPDs of each node given their parents.

Proof in general case?

Example: Markov Chain

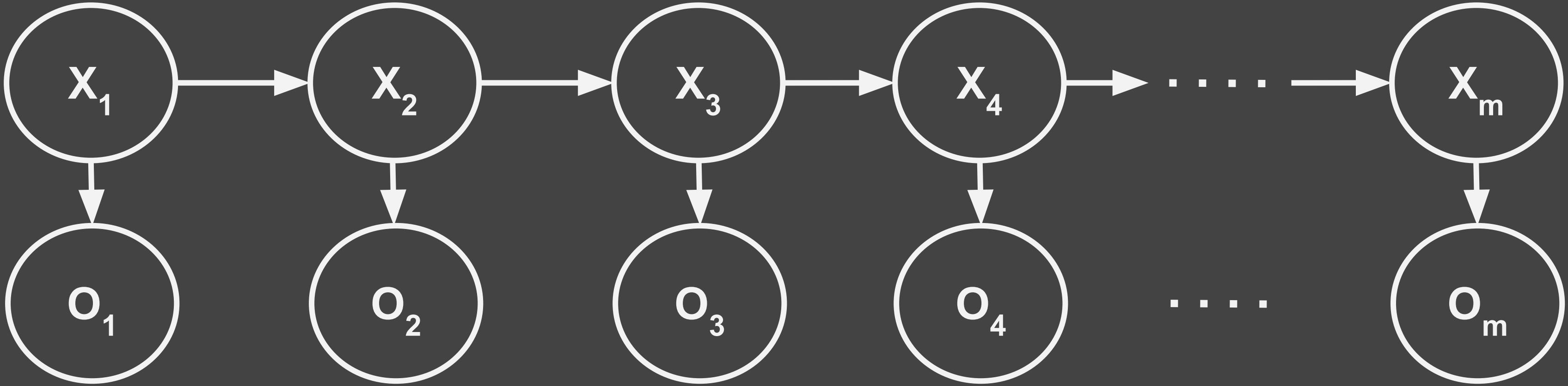


$$p(x_t \mid x_{t-1}, x_{t-2}, \dots, x_2, x_1) = p(x_t \mid x_{t-1})$$

$$p(x_1, x_2, \dots, x_m) = \dots$$

$$= p(x_1) \;\; p(x_2|x_1) \;\; p(x_3|x_2) \;\; \dots \;\; p(x_{m-1}|x_{m-2}) \;\; p(x_m|x_{m-1})$$

Example: Hidden Markov Models, Bayesian Filters

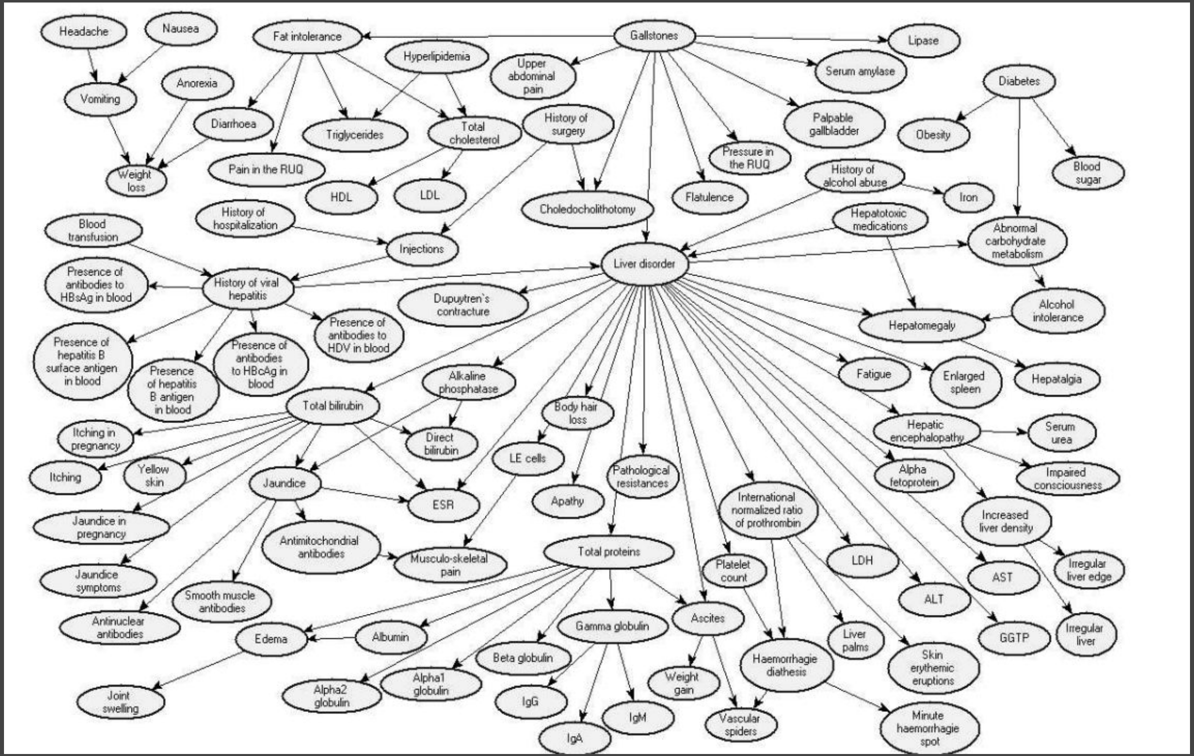


$$P(x_1, x_2, \ldots, x_m, o_1, o_2, \ldots, o_m) = \ldots$$

$$= p(x_1)\ p(x_2|x_1)\ p(x_3|x_2) \ldots p(x_{m-1}|x_{m-2})\ p(x_m|x_{m-1})$$

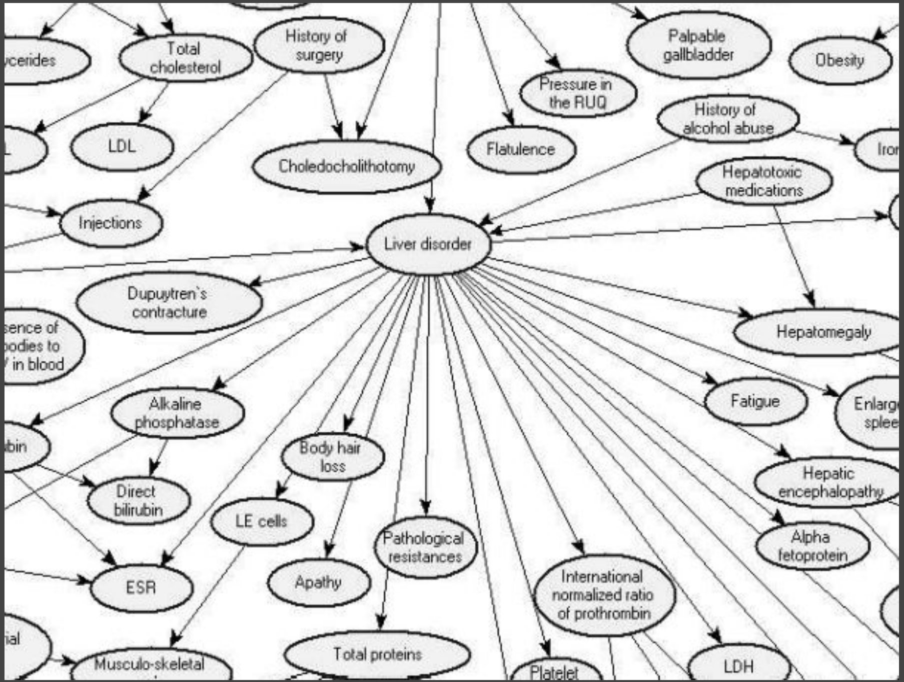
$$p(o_1|x_1)\ p(o_2|x_2) \ldots p(o_{m-1}|x_{m-1})\ p(o_m|x_m)$$

Example: Medical Diagnosis



Onisko, Agnieszka, et al. "A Bayesian network model for diagnosis of liver disorders.", 1999.

Example: Medical Diagnosis



Onisko, Agnieszka, et al. "A Bayesian network model for diagnosis of liver disorders.", 1999.

Bayesian Nets and Causality

K. N. Toosi
University of Technology

$$P(C,R,W) = P(W \mid R)\, P(R \mid C)\, P(C)$$

$$P(C,R,W) = P(C \mid R,W)\, P(R,W) = P(C \mid R)\, P(R \mid W)\, P(W)$$

Bayesian Nets and Causality

$$P(C,R,W) = P(W \mid R)\, P(R|C)\, P(C)$$

$$P(C,R,W) = P(C \mid R,W)\, P(R,W) = P(C \mid R)\, P(R \mid W)\, P(W)$$

Bayesian Nets and Causality

$$P(C,R,W) = P(W \mid R)\, P(R \mid C)\, P(C)$$

$$P(C,R,W) = P(C,W \mid R)\, P(R) = P(C \mid R)\, P(W \mid R)\, P(R)$$

Bayesian Nets and Causality

What about ....

K. N. Toosi
University of Technology