

Universität Potsdam  
Institut für Informatik  
Lehrstuhl Maschinelles Lernen

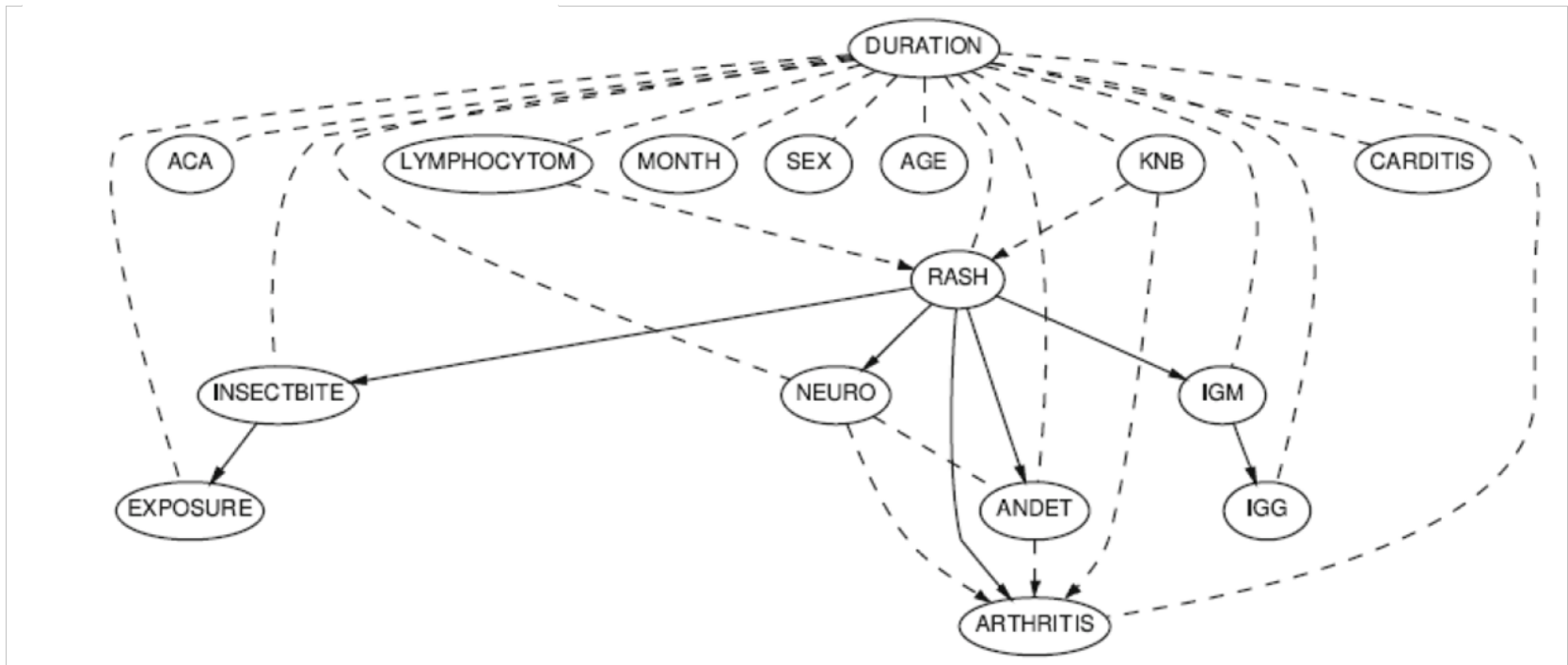


# Graphische Modelle

Christoph Sawade/Niels Landwehr/Tobias Scheffer

# Graphische Modelle

- Werkzeug zur Modellierung einer Domäne mit verschiedenen Zufallsgrößen
- Gemeinsame Verteilung, insb. Abhängigkeiten



# Überblick

- Gerichtete Graphische Modelle: Bayessche Netze
- Graphische Modelle im Maschinellen Lernen
- Inferenz in Graphischen Modellen
- Ungerichtete Graphische Modelle: Markov Netze

# Überblick

- Gerichtetete Graphische Modelle: Bayessche Netze
- Graphische Modelle im Maschinellen Lernen
- Inferenz in Graphischen Modellen
- Ungerichtete Graphische Modelle: Markov Netze

# Zufallsvariablen, Verteilungen

- Zufallsvariable: Abbildung eines Ereignisses auf reelle Zahl

$$X : \Omega \rightarrow \mathbb{R} \quad \Omega \text{ Raum der Elementarereignisse}$$

- Verteilung über ZV: wie wahrscheinlich, bestimmte Werte zu beobachten?

- Diskrete Zufallsvariablen: diskreter Wertebereich (z.B. Münzwurf)

Diskrete Wahrscheinlichkeit  $p(x) \in [0,1]$ ,  $\sum_x p(x) = 1$   $x$  Wert der ZV

- Kontinuierliche Zufallsvariablen: kontinuierlicher Wertebereich (z.B. Körpergröße)

Dichtefunktion  $p(x) \in \mathbb{R}_{\geq 0}$ ,  $\int_{-\infty}^{\infty} p(x) dx = 1$   $x$  Wert der ZV

# Zufallsvariablen, Verteilungen

- Beispiele für diskrete/kontinuierliche Verteilungen

- ◆ Bernoulli-Verteilung: binäre ZV  $X \in \{0,1\}$

Parameter  $\mu = p(X = 1 | \mu)$

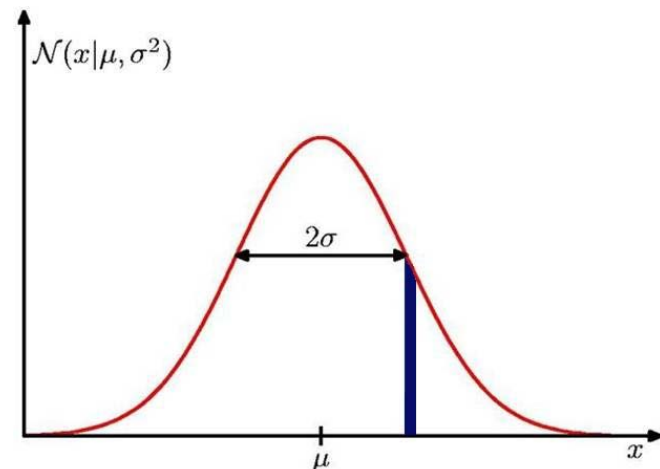
$$X \sim \text{Bern}(X | \mu) = \mu^X (1 - \mu)^{1-X}$$

- ◆ Normalverteilung: kontinuierliche ZV  $X \in \mathbb{R}$

$$N(x | \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$$

$\mu$  Mittelwert

$\sigma$  Standardabweichung

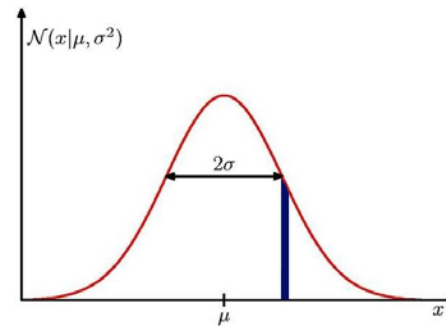


# Zufallsvariablen, Verteilungen

- Zusammenhang Dichte/Wahrscheinlichkeit

$$p(X \in [x - \varepsilon, x + \varepsilon]) = \int_{x - \varepsilon}^{x + \varepsilon} p(z) dz$$

$p(X \in [x - \varepsilon, x + \varepsilon])$  ist  
(diskrete) Wahrscheinlichkeit



- Oft abstrahieren wir vom Typ der Zufallsvariable
  - ◆ Rechenregeln sind im Wesentlichen gleich für diskrete/kontinuierliche Variablen („ersetze Summe durch Integral“)
  - ◆ Grundlegende Begriffe wie Abhängigkeit, Erwartungswert, Varianz sind auf beide Typen von Variablen anwendbar

# Gemeinsame Verteilung, bedingte Verteilung

- Gemeinsame Verteilung  $p(X, Y)$  über ZV  $X, Y$

$$p(x, y) \in [0, 1], \quad \sum_{x, y} p(x, y) = 1 \quad p(x, y) \in \mathbb{R}_{\geq 0}, \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) dx dy = 1$$

- Bedingte Verteilung

$$p(X | Y) = \frac{p(X, Y)}{p(Y)} \quad \text{diskret oder kontinuierlich}$$

Für  $p(y) > 0$  ist  $p(X | y)$  wieder eine Verteilung über  $X$ :

$$\sum_x p(x | y) = 1 \quad \text{für } p(y) > 0 \quad (\text{diskret})$$

$$\int_{-\infty}^{\infty} p(x | y) dx = 1 \quad \text{für } p(y) > 0 \quad (\text{kontinuierlich})$$



# Unabhängigkeit von Zufallsvariablen

- Unabhängigkeit (diskret oder kontinuierlich)

X,Y unabhängig genau dann wenn  $p(X, Y) = p(X)p(Y)$

X,Y unabhängig genau dann wenn  $p(X | Y) = p(X)$

X,Y unabhängig genau dann wenn  $p(Y | X) = p(Y)$

- Bedingte Unabhängigkeit (diskret oder kontinuierlich)

X,Y unabhängig gegeben Z genau dann wenn  $p(X, Y | Z) = p(X | Z)p(Y | Z)$

X,Y unabhängig gegeben Z genau dann wenn  $p(Y | X, Z) = p(Y | Z)$

X,Y unabhängig gegeben Z genau dann wenn  $p(X | Y, Z) = p(X | Z)$

... einfach Anwendung des Unabhängigkeitsbegriffs auf die bedingte gemeinsame Verteilung  $p(X, Y|Z)$

# Rechenregeln Verteilungen

- Rechenregeln Wahrscheinlichkeiten/Verteilungen

- ◆ Summenregel

$$p(x) = \sum_y p(x, y) \quad \text{diskrete Verteilungen}$$

$$p(x) = \int_{-\infty}^{\infty} p(x, y) dy \quad \text{kontinuierliche Verteilungen}$$

$p(X)$  heisst auch Randverteilung, Marginalverteilung

- ◆ Produktregel

$$p(X, Y) = p(X | Y)p(Y) \quad \text{diskret oder kontinuierlich}$$

- Wahrscheinlichkeitsrechnungen beruhen auf Anwendungen dieser beiden Regeln

# Graphische Modelle: Idee/Ziel

- Ziel: Modellierung der gemeinsamen Verteilung  $p(X_1, \dots, X_N)$  einer Menge von ZV  $X_1, \dots, X_N$
- Aus  $p(X_1, \dots, X_N)$  lassen sich berechnen...
  - ◆ Alle Randverteilungen (Summenregel)

$$p(X_{i_1}, \dots, X_{i_m}), \quad \{i_1, \dots, i_m\} \subseteq \{1, \dots, N\}$$

- ◆ Alle bedingten Verteilungen (aus Randverteilungen)

$$p(X_{i_1}, \dots, X_{i_m} \mid X_{i_{m+1}}, \dots, X_{i_{m+k}}), \quad \{i_1, \dots, i_{m+k}\} \subseteq \{1, \dots, N\}$$

- Damit lassen sich alle probabilistischen Fragestellungen (*Inferenzprobleme*) über  $X_1, \dots, X_N$  beantworten

# Graphische Modelle: Idee/Ziel

- Graphische Modelle: Kombination von Wahrscheinlichkeitstheorie und Graphentheorie
- Kompakte, intuitive Modellierung von  $p(X_1, \dots, X_N)$ 
  - ◆ Graphstruktur repräsentiert Abhängigkeiten zwischen Variablen  $X_1, \dots, X_N$
  - ◆ Einsicht in Struktur des Modells; einfach, Vorwissen einzubringen
  - ◆ Effiziente Algorithmen für Inferenz, die Graphstruktur ausnutzen
- Viele Methoden des maschinellen Lernens lassen sich in Form von GM darstellen
- Fragestellungen wie MAP Lernen, Bayessche Vorhersage lassen sich als Inferenzprobleme in GM formulieren

# Graphische Modelle: Beispiel

- Beispiel: „Alarm“ Szenario
  - ◆ Unser Haus in LA hat eine Alarmanlage.
  - ◆ Wir sind im Urlaub. Unser Nachbar ruft an, falls er den Alarm hört. Wenn eingebrochen wurde, wollen wir zurück kommen.
  - ◆ Leider ist der Nachbar nicht immer zu Hause.
  - ◆ Leider geht die Alarmanlage auch bei kleinen Erdbeben los.
- 5 binäre Zufallsvariablen
  - ⓑ Burglary – Einbruch hat stattgefunden
  - ⓔ Earthquake – Erdbeben hat stattgefunden
  - ⓐ Alarm – Alarmanlage geht los
  - Ⓝ NeighborCalls – Nachbar ruft an
  - Ⓡ RadioReport – Bericht über Erdbeben im Radio

# Graphische Modelle: Beispiel

- Zufallsvariablen haben eine gemeinsame Verteilung  $p(B, E, A, N, R)$ . Wie angeben? Welche Abhängigkeiten gelten?
- Beispiel für Inferenzproblem: Nachbar hat angerufen ( $N=1$ ), wie wahrscheinlich ist Einbruch ( $B=1$ )?
  - ◆ Hängt von verschiedenen Faktoren ab
    - ★ Wie wahrscheinlich Einbruch a priori?
    - ★ Wie wahrscheinlich Erdbeben a priori?
    - ★ Wie wahrscheinlich, dass Alarmanlage auslöst?
    - ★ ...

$$\begin{aligned} \text{(Naive) Inferenz: } p(B=1 | N=1) &= \frac{p(B=1, N=1)}{p(N=1)} \\ &= \frac{\sum_E \sum_A \sum_R p(B=1, E, A, N=1, R)}{\sum_B \sum_E \sum_A \sum_R p(B, E, A, N=1, R)} \end{aligned}$$

# Graphische Modelle: Beispiel

- Wie modellieren wir  $p(B, E, A, N, R)$ ?
  - ◆ 1. Versuch: vollständige Tabelle

$2^N$  {

$B$	$E$	$A$	$N$	$R$	$P(B, E, A, N, R)$
0	0	0	0	0	0.6
1	0	0	0	0	0.005
0	1	0	0	0	0.01
...	...	...	...	...	...

- + Alle Verteilungen  $p(B, E, A, N, R)$  können repräsentiert werden
- Anzahl Parameter exponentiell
- Schwierig anzugeben

- ◆ 2. Versuch: alles unabhängig

$$p(B, E, A, N, R) = p(B)p(E)p(A)p(N)p(R)$$

+ Anzahl Parameter linear

- Zu restriktiv, Unabhängigkeitsannahme erlaubt keine sinnvolle Inferenz

# Graphische Modelle: Beispiel

- Graphisches Modell: Selektive Unabhängigkeitsannahmen, durch Vorwissen motiviert
- Wähle Variablenordnung: z.B.  $B < E < A < N < R$
- Produktregel:

$$\begin{aligned} p(B, E, A, N, R) &= p(B, E, A, N) p(R | B, E, A, N) \\ &= p(B, E, A) p(N | B, E, A) p(R | B, E, A, N) \\ &= p(B, E) p(A | B, E) p(N | B, E, A) p(R | B, E, A, N) \\ &= p(B) p(E | B) p(A | B, E) p(N | B, E, A) p(R | B, E, A, N) \end{aligned}$$

Faktoren beschreiben die Verteilung einer Zufallsvariablen in Abhängigkeit anderer Zufallsvariablen.

Können wir diese Faktoren vereinfachen?

Welche dieser Abhängigkeiten bestehen wirklich?



# Graphische Modelle: Beispiel

- Zerlegung in Faktoren nach Produktregel:

$$p(B, E, A, N, R) = p(B)p(E | B)p(A | B, E)p(N | B, E, A)p(R | B, E, A, N)$$

- Annahme bedingter Unabhängigkeiten (Entfernen von Variablen aus Bedingungsteil)

$$p(E | B) = p(E)$$

*Erdbeben hängt nicht von Einbruch ab*

$$p(A | B, E) = p(A | B, E)$$

*Alarm hängt von Einbruch und Erdbeben ab*

$$p(N | B, E, A) = p(N | A)$$

*Anruf von Nachbar hängt nur von Alarm ab*

$$p(R | B, E, A, N) = p(R | E)$$

*Nachricht im Radio hängt nur Erdbeben ab*

- Vereinfachte Darstellung der gemeinsamen Verteilung:

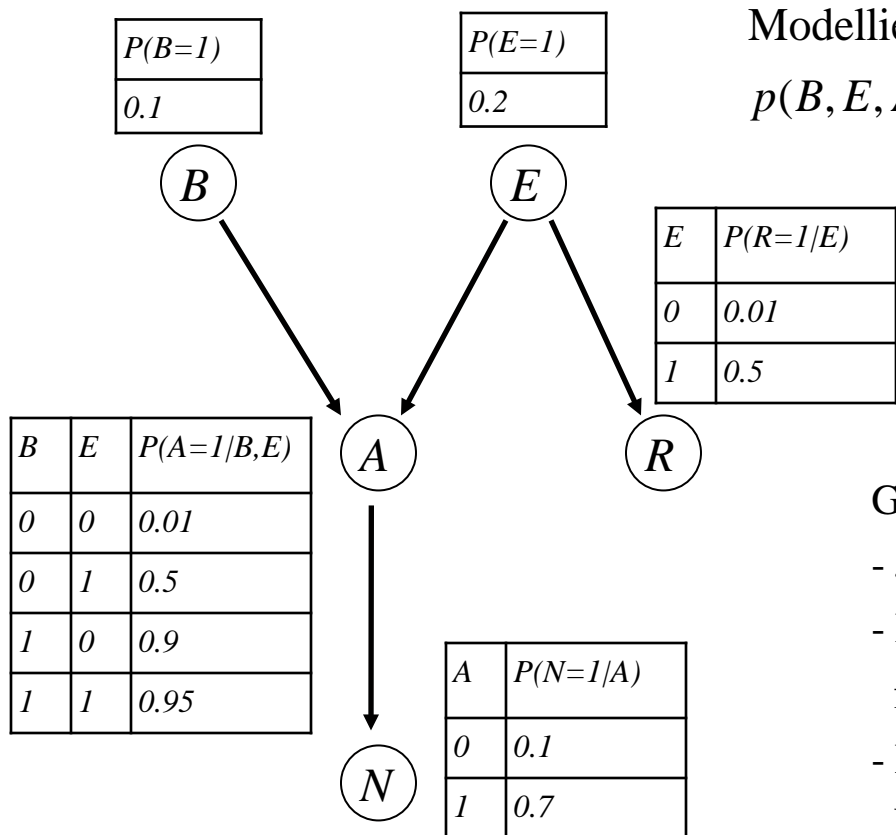
$$p(B, E, A, N, R) = p(B)p(E)p(A | E, B)p(N | A)p(R | E)$$

Vereinfachte Faktoren



# Graphische Modelle: Beispiel

## ■ Graphisches Modell für „Alarm“ Szenario



Modellierte Verteilung:

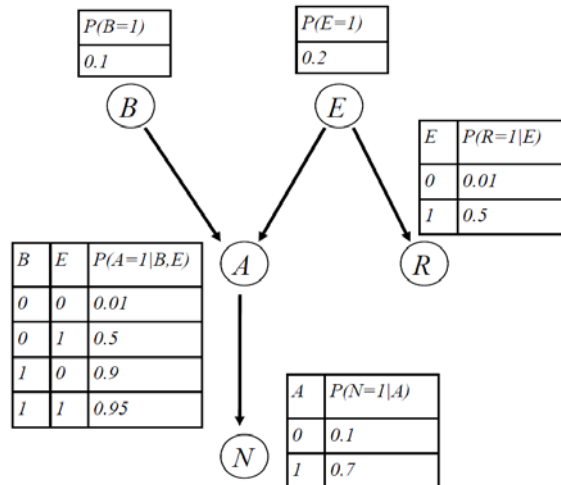
$$p(B, E, A, N, R) = p(B)p(E)p(A | E, B)p(N | A)p(R | E)$$

Graphisches Modell:

- Jede ZV ist ein Knoten
- Für jeden Faktor der Form  $p(X | X_1, \dots, X_k)$  fügen wir gerichtete Kanten von den  $X_i$  zu  $X$  ein
- Modell ist parametrisiert mit den bedingten Verteilungen  $p(X | X_1, \dots, X_k)$

# Graphische Modelle: Beispiel

- Graphisches Modell für „Alarm“ Szenario



- ◆ Anzahl Parameter:  $O(N*2^K)$ ,  $K$  = max. Anzahl von Elternknoten
- ◆ Hier 1+1+2+2+4 statt  $2^5-1$  Parameter
- Gerichtete graphische Modelle heißen auch **Bayessche Netze**

# Gerichtete Graphische Modelle: Definition

- Gegeben eine Menge von ZV  $\{X_1, \dots, X_N\}$
- Ein gerichtetes graphisches Modell über den ZV  $\{X_1, \dots, X_N\}$  ist ein gerichteter Graph mit
  - ◆ Knotenmenge  $X_1, \dots, X_N$
  - ◆ Es gibt keine gerichteten Zyklen  $X_{i_1} \rightarrow X_{i_2} \rightarrow \dots \rightarrow X_{i_k} \rightarrow X_{i_1}$
  - ◆ Knoten sind mit parametrisierten bedingten Verteilungen  $p(X_i | pa(X_i))$  assoziiert, wobei  $pa(X_i) = \{X_j | X_j \rightarrow X_i\}$  die Menge der Elternknoten eines Knoten ist
- Das graphische Modell modelliert eine gemeinsame Verteilung über  $X_1, \dots, X_N$  durch

$$p(X_1, \dots, X_N) = \prod_{i=1}^N p(X_i | pa(X_i))$$

# Gerichtete Graphische Modelle: Definition

- Warum muss der Graph azyklisch sein?

- ◆ Satz aus der Graphentheorie:

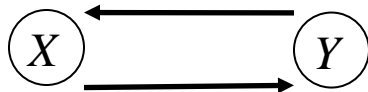
$G$  ist azyklisch  $\Leftrightarrow$  es gibt Ordnung  $\leq_G$  der Knoten, so dass gerichtete Kanten die Ordnung respektieren ( $N \rightarrow N' \Rightarrow N \leq_G N'$ )

- ◆ Damit ergibt sich

$$p(X_1, \dots, X_N) = \prod_{i=1}^N p(X_i \mid pa(X_i))$$

aus Produktregel + bedingten Unabhängigkeitsannahmen  
(Variablen entsprechend  $\leq_G$  umsordieren)

- Gegenbeispiel (kein graphisches Modell):



$$p(X, Y) \neq p(X \mid Y)p(Y \mid X)$$

# Graphische Modelle: Unabhängigkeit

- Die Graphstruktur eines Graphischen Modells impliziert (bedingte) Unabhängigkeiten zwischen ZV
- Notation: Für Variablen  $X, Y, Z$  schreiben wir

$$X \perp Y | Z \Leftrightarrow p(X | Y, Z) = p(X | Z)$$

" $X$  unabhängig von  $Y$  gegeben  $Z$ "

- Erweiterung auf disjunkte Mengen  $A, B, C$  von ZV

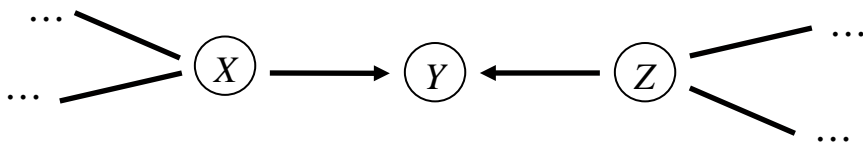
$$A \perp B | C \Leftrightarrow p(A | B, C) = p(A | C)$$

# Graphische Modelle: Unabhängigkeit

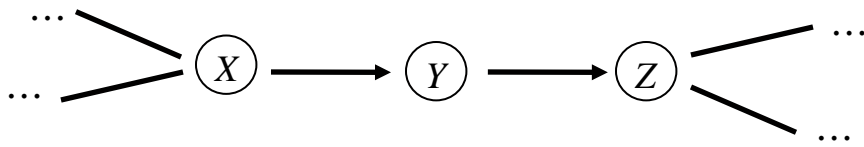
- Welche Unabhängigkeiten der Form  $A \perp B | C$  werden durch die Graphstruktur modelliert?
  - ◆ Im Prinzip auszurechnen durch Summen/Produktregel aus der gemeinsamen Verteilung
  - ◆ Bei graphischen Modellen aber direkt aus der Graphstruktur ableitbar → viel einfacher
  - ◆ „D-separation“: Menge einfacher Regeln, nach denen sich alle Unabhängigkeiten ableiten lassen

# Graphische Modelle: Unabhängigkeit

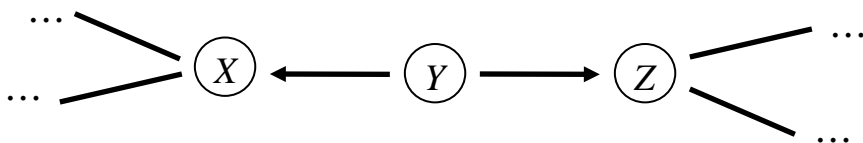
- D-separation: Welche Unabhängigkeiten der Form  $A \perp B | C$  werden durch die Graphstruktur modelliert?
- Wichtige Rolle spielen Pfade im Graphen, die ZV verbinden
- Notation:



Pfad zwischen X und Z hat eine „**konvergierende**“ Verbindung bei Y („head to head“)



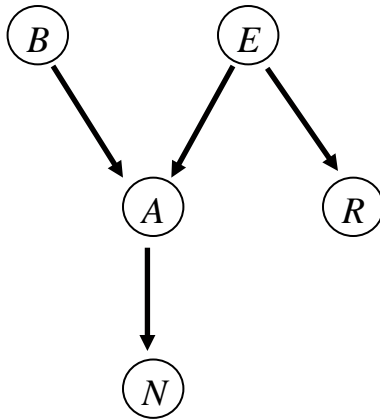
Pfad zwischen X und Z hat eine „**serielle**“ Verbindung bei Y („head to tail“)



Pfad zwischen X und Z hat eine „**divergierende**“ Verbindung bei Y („tail-to-tail“)



# Divergierende Verbindungen



Gemeinsame Verteilung:

$$p(B, E, A, N, R) =$$

$$p(B)p(E)p(A|E, B)p(N|A)p(R|E)$$

$B$  = „Einbruch“

$E$  = „Erdbeben“

$A$  = „Alarm“

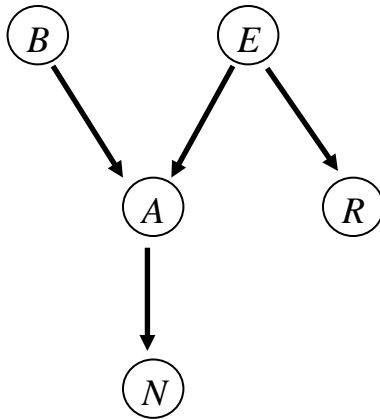
$N$  = „Nachbar ruft an“

$R$  = „Radio Bericht“

- Betrachte Pfad  $A \leftarrow E \rightarrow R$ . Gilt  $A \perp R | \emptyset$  ?



# Divergierende Verbindungen



Gemeinsame Verteilung:

$$p(B, E, A, N, R) =$$

$$p(B)p(E)p(A | E, B)p(N | A)p(R | E)$$

$B$  = „Einbruch“

$N$  = „Nachbar ruft an“

$E$  = „Erdbeben“

$R$  = „Radio Bericht“

$A$  = „Alarm“

## ■ Betrachte Pfad $A \leftarrow E \rightarrow R$ . Gilt $A \perp R | \emptyset$ ?

◆ Nein,  $p(A | R) \neq p(A)$  [Ausrechnen mit gemeinsamer Verteilung]

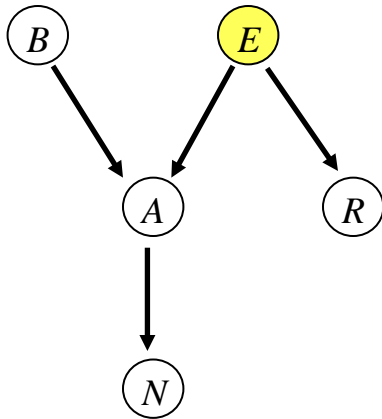
◆ Intuitiv:

RadioReport  $\Rightarrow$  wahrscheinlich Erdbeben  $\Rightarrow$  wahrscheinlich Alarm

$$p(A = 1 | R = 1) > p(A = 1 | R = 0)$$

◆ ZV  $R$  beeinflusst ZV  $A$  über die divergierende Verbindung  $R \leftarrow E \rightarrow A$

# Divergierende Verbindungen



Gemeinsame Verteilung:

$$p(B, E, A, N, R) =$$

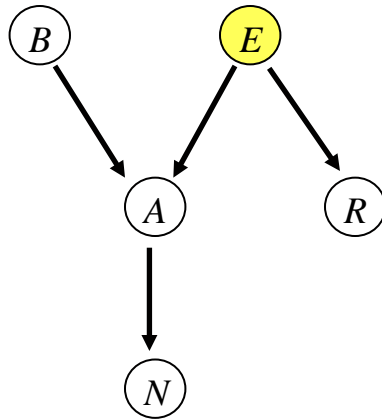
$$p(B)p(E)p(A | E, B)p(N | A)p(R | E)$$

 beobachteter Knoten

- Betrachte Pfad  $A \leftarrow E \rightarrow R$ . Gilt  $A \perp R | E$  ?



# Divergierende Verbindungen



Gemeinsame Verteilung:

$$p(B, E, A, N, R) =$$

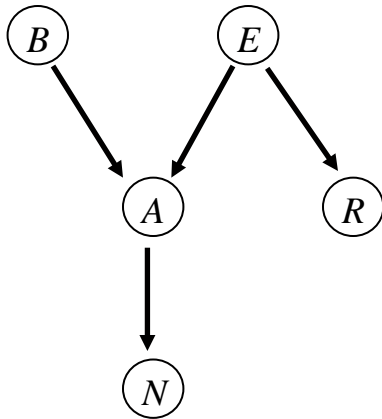
$$p(B)p(E)p(A | E, B)p(N | A)p(R | E)$$

 beobachteter Knoten

## ■ Betrachte Pfad $A \leftarrow E \rightarrow R$ . Gilt $A \perp R | E$ ?

- ◆ Ja,  $p(A | R, E) = p(A | E)$  [Ausrechnen mit gemeinsamer Verteilung]
- ◆ Intuitiv:  
Wenn wir schon wissen, dass ein Erdbeben eingetreten ist, wird die Wahrscheinlichkeit für Alarm nicht höher/niedriger durch RadioReport
- ◆ Der divergierende Pfad  $R \leftarrow E \rightarrow A$  wird durch Beobachtung von E blockiert

# Serielle Verbindungen



Gemeinsame Verteilung:

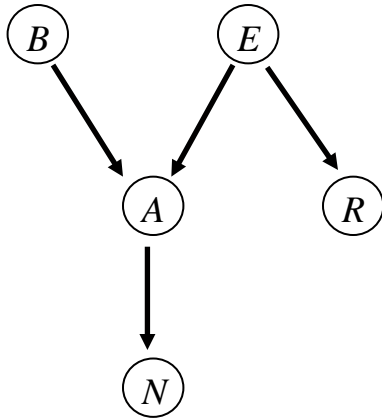
$$p(B, E, A, N, R) =$$

$$p(B)p(E)p(A|E, B)p(N|A)p(R|E)$$

- Betrachte Pfad  $N \leftarrow A \leftarrow B$ . Gilt  $B \perp N | \emptyset$  ?



# Serielle Verbindungen



Gemeinsame Verteilung:

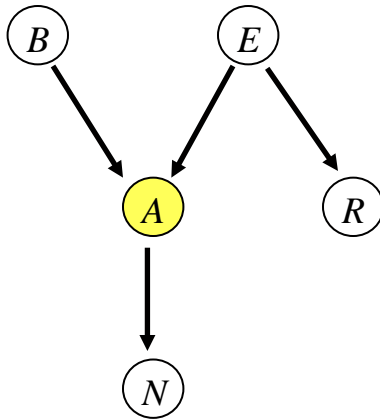
$$p(B, E, A, N, R) =$$

$$p(B)p(E)p(A|E, B)p(N|A)p(R|E)$$

■ **Betrachte Pfad  $N \leftarrow A \leftarrow B$ . Gilt  $B \perp N | \emptyset$  ?**

- ◆ Nein,  $p(B|N) \neq p(B)$  [Ausrechnen mit gemeinsamer Verteilung]
- ◆ Intuitiv:  
NeighborCalls  $\Rightarrow$  wahrscheinlich Alarm  $\Rightarrow$  wahrscheinlich Burglary  
 $p(B=1|N=1) > p(B=1|N=0)$
- ◆ ZV  $N$  beeinflusst ZV  $B$  über den seriellen Pfad  $N \leftarrow A \leftarrow B$

# Serielle Verbindungen



Gemeinsame Verteilung:

$$p(B, E, A, N, R) =$$

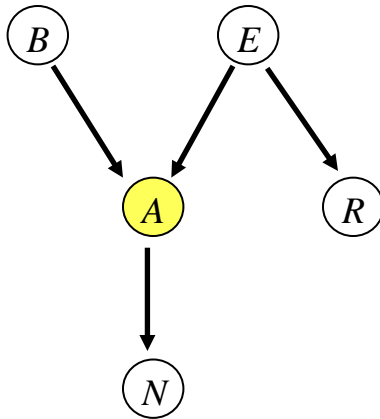
$$p(B)p(E)p(A|E, B)p(N|A)p(R|E)$$

 beobachteter Knoten

- Betrachte Pfad  $N \leftarrow A \leftarrow B$ . Gilt  $B \perp N | A$  ?



# Serielle Verbindungen



Gemeinsame Verteilung:

$$p(B, E, A, N, R) =$$

$$p(B)p(E)p(A | E, B)p(N | A)p(R | E)$$

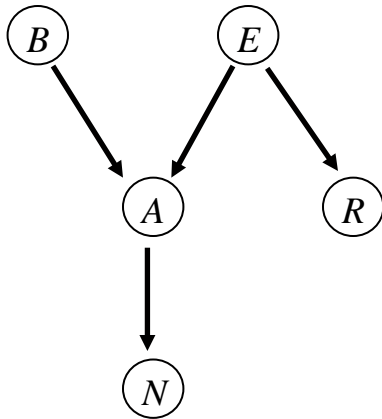
 beobachteter Knoten

## ■ Betrachte Pfad $N \leftarrow A \leftarrow B$ . Gilt $B \perp N | A$ ?

- ◆ Ja,  $p(B | N, A) = p(B | A)$  [Ausrechnen mit gemeinsamer Verteilung]
- ◆ Intuitiv:  
Wenn wir schon wissen, dass der Alarm ausgelöst wurde, sinkt/steigt die Wahrscheinlichkeit für Einbruch nicht dadurch, dass Nachbar anruft
- ◆ Der serielle Pfad  $N \leftarrow A \leftarrow B$  wird durch Beobachtung von  $A$  blockiert.



# Konvergierende Verbindung



Gemeinsame Verteilung:

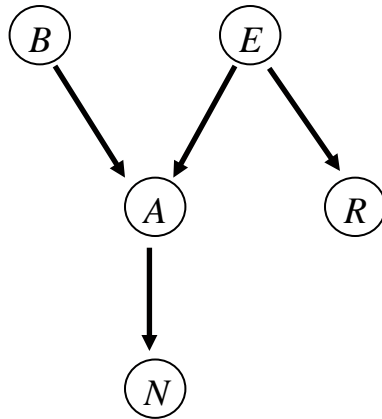
$$p(B, E, A, N, R) =$$

$$p(B)p(E)p(A|E, B)p(N|A)p(R|E)$$

- Betrachte Pfad  $B \rightarrow A \leftarrow E$ . Gilt  $B \perp E | \emptyset$  ?



# Konvergierende Verbindung



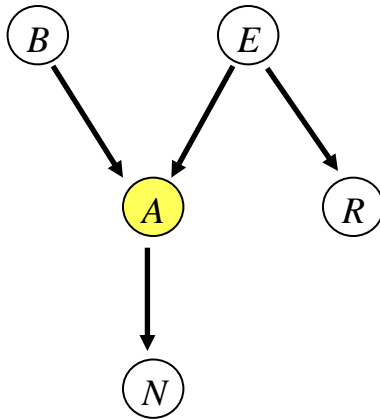
Gemeinsame Verteilung:

$$p(B, E, A, N, R) =$$

$$p(B)p(E)p(A | E, B)p(N | A)p(R | E)$$

- Betrachte Pfad  $B \rightarrow A \leftarrow E$ . Gilt  $B \perp E | \emptyset$  ?
  - ◆ Ja,  $p(B | E) = p(B)$  [Ausrechnen mit gemeinsamer Verteilung]
  - ◆ Intuitiv:  
Einbrüche treten nicht häufiger/seltener auf an Tagen mit Erdbeben
  - ◆ Der konvergierende Pfad  $B \rightarrow A \leftarrow E$  ist blockiert wenn  $A$  **nicht** beobachtet ist

# Konvergierende Verbindung



Gemeinsame Verteilung:

$$p(B, E, A, N, R) =$$

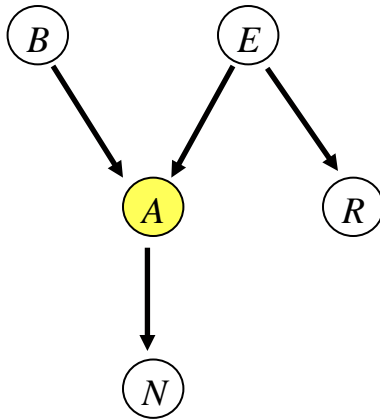
$$p(B)p(E)p(A|E, B)p(N|A)p(R|E)$$

 beobachteter Knoten

- Betrachte Pfad  $B \rightarrow A \leftarrow E$ . Gilt  $B \perp E | A$  ?



# Konvergierende Verbindung



Gemeinsame Verteilung:

$$p(B, E, A, N, R) =$$

$$p(B)p(E)p(A | E, B)p(N | A)p(R | E)$$

 beobachteter Knoten

## ■ Betrachte Pfad $B \rightarrow A \leftarrow E$ . Gilt $B \perp E | A$ ?

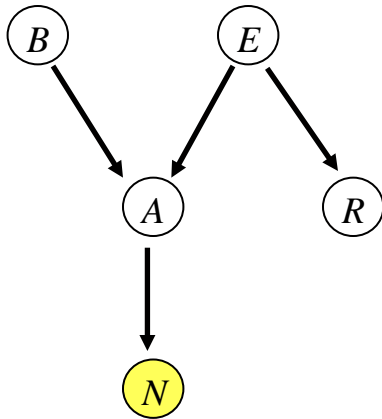
◆ Nein,  $p(B | E, A) \neq p(B | A)$  [Ausrechnen mit gemeinsamer Verteilung]

◆ Intuitiv:

Alarm wurde ausgelöst. Falls wir ein Erdbeben beobachten, erklärt das den Alarm, Wahrscheinlichkeit für Einbruch sinkt ("explaining away").

◆ Der konvergierende Pfad  $B \rightarrow A \leftarrow E$  wird **freigegeben** durch Beobachtung von A

# Konvergierende Verbindung



Gemeinsame Verteilung:

$$p(B, E, A, N, R) =$$

$$p(B)p(E)p(A | E, B)p(N | A)p(R | E)$$

 beobachteter Knoten

## ■ Betrachte Pfad $B \rightarrow A \leftarrow E$ . Gilt $B \perp E | N$ ?

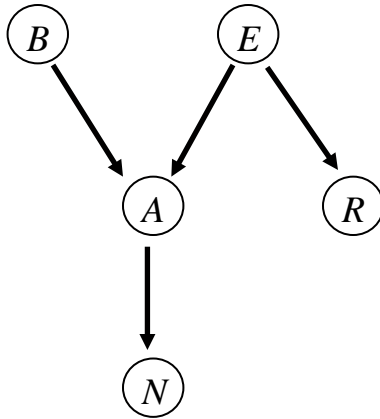
◆ Nein,  $p(B | N, A) \neq p(B | A)$  [Ausrechnen mit gemeinsamer Verteilung]

◆ Intuitiv:

NeighborCalls indirekte Beobachtung von Alarm. Erdbebenbeobachtung erklärt den Alarm, Wahrscheinlichkeit für Einbruch sinkt ("explaining away").

◆ Der konvergierende Pfad  $B \rightarrow A \leftarrow E$  wird **freigegeben** durch Beobachtung von N

# Zusammenfassung Pfade



Gemeinsame Verteilung:

$$p(B, E, A, N, R) =$$

$$p(B)p(E)p(A | E, B)p(N | A)p(R | E)$$

- Zusammenfassung: ein Pfad  $\dots-X-Y-Z-\dots$  ist
  - ◆ Blockiert bei  $Y$ , wenn
    - ★ Divergierende Verbindung, und  $Y$  beobachtet, oder
    - ★ Serielle Verbindung, und  $Y$  beobachtet, oder
    - ★ Konvergierende Verbindung, und weder  $Y$  noch einer seiner Nachfahren  $Y' \in \text{Descendants}(Y)$  beobachtet
    - ★  $\text{Descendants}(Y) = \{Y' \mid \text{es gibt gerichteten Pfad von } Y \text{ zu } Y'\}$
  - ◆ Sonst ist der Pfad frei bei  $Y$

# D-Separation: Blockierte Pfade

- Ein Pfad ist blockiert, wenn er bei einem auf dem Pfad liegenden Knoten blockiert ist:
  - ◆ Seien  $X, X' \in ZV$ ,  $C$  eine beobachtete Menge von  $ZV$ ,  $X, X' \notin C$
  - ◆ Ein ungerichteter Pfad  $X - X_1 - \dots - X_n - X'$  zwischen  $X$  und  $X'$  ist blockiert gegeben  $C$  gdw es einen Knoten  $X_i$  gibt so dass Pfad bei  $X_i$  blockiert ist gegeben  $C$
- D-Separation basiert auf blockierten Pfaden:
  - ◆ Seien  $A, B, C$  disjunkte Mengen von  $ZV$ .
  - ◆ Definition:  $A$  und  $B$  sind d-separiert durch  $C$  gdw jeder Pfad von einem Knoten  $X \in A$  zu einem Knoten  $Y \in B$  blockiert ist gegeben  $C$ .

# D-Separation: Korrektheit

- Gegeben ein graphisches Modell über  $\{X_1, \dots, X_N\}$  mit Graphstruktur  $G$ .
- Das graphische Modell modelliert eine Verteilung durch

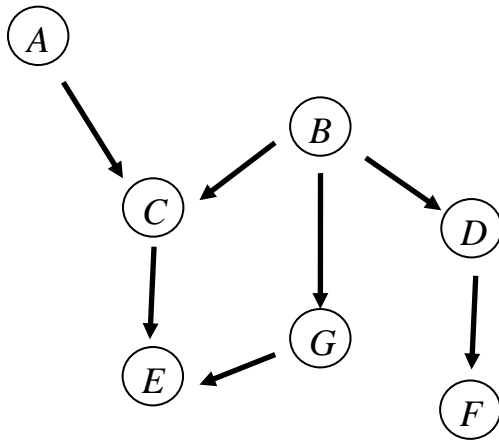
$$p(X_1, \dots, X_N) = \prod_{i=1}^N p(X_i | pa(X_i))$$

abhängig von den bedingten Verteilungen  $p(X_n | pa(X_n))$ .

- Theorem (Korrektheit, Vollständigkeit d-separation)
  - ◆ Falls  $A, B$  d-separiert gegeben  $C$  in  $G$ , dann  $p(A | B, C) = p(A | C)$
  - ◆ Es gibt keine anderen Unabhängigkeiten, die für jede Wahl der bedingten Verteilungen  $p(X_i | pa(X_i))$  gelten.



# D-Separation: Beispiel



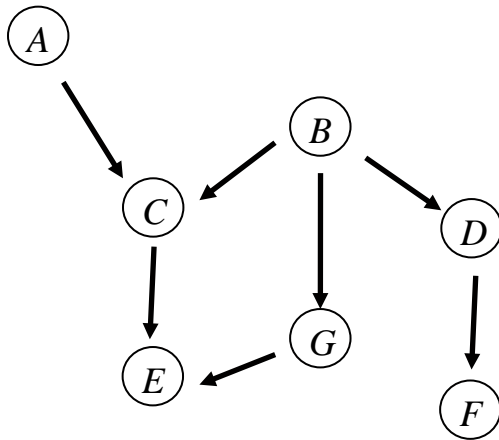
Gilt  $A \perp F \mid D$ ?

Gilt  $B \perp E \mid C$ ?

Gilt  $A \perp E \mid C$ ?

- Ein Pfad  $\dots-X-Y-Z-\dots$  ist
  - ◆ Blockiert bei  $Y$ , wenn
    - ★ Divergierende Verbindung, und  $Y$  beobachtet, oder
    - ★ Serielle Verbindung, und  $Y$  beobachtet, oder
    - ★ Konvergierende Verbindung, und weder  $Y$  noch einer seiner Nachfahren  $Y' \in \text{Descendants}(Y)$  beobachtet
  - ◆ Sonst ist der Pfad frei bei  $Y$

# D-Separation: Beispiel



Gilt  $A \perp F \mid D$ ?

Ja

Gilt  $B \perp E \mid C$ ?

Nein:  $B - G - E$

Gilt  $A \perp E \mid C$ ?

Nein:  $A - C - B - G - E$

- Ein Pfad  $\dots-X-Y-Z-\dots$  ist
  - ◆ Blockiert bei  $Y$ , wenn
    - ★ Divergierende Verbindung, und  $Y$  beobachtet, oder
    - ★ Serielle Verbindung, und  $Y$  beobachtet, oder
    - ★ Konvergierende Verbindung, und weder  $Y$  noch einer seiner Nachfahren  $Y' \in \text{Descendants}(Y)$  beobachtet
  - ◆ Sonst ist der Pfad frei bei  $Y$

# Graphische Modelle: Kausalität

- Oft werden Modelle so konstruiert, dass gerichtete Kanten kausalen Einflüssen entsprechen



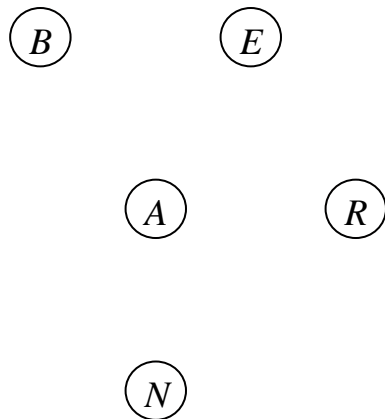
- Äquivalentes Modell



- **Definition:**  $I(G) = \{ (A \perp B \mid C) : A \text{ und } B \text{ sind d-separiert gegeben } C \text{ in } G \}$   
„Alle Unabhängigkeitsannahmen, die durch  $G$  kodiert werden“
- $I(G) = I(G') = \emptyset$  :
  - ◆ Keine statistischen Gründe, eines der Modelle vorzuziehen
  - ◆ Kann nicht aufgrund von Daten zwischen Modellen unterscheiden
  - ◆ Aber „kausale“ Modelle oft besser verständlich

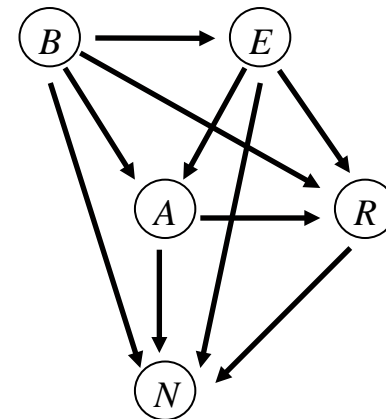
# Modelle Unterschiedlicher Komplexität

- Komplexität: bestimmt durch Menge der Kanten im Graph
  - ◆ Beeinflusst die Anzahl der Parameter im Modell. Für binäre ZV gilt: Anzahl Parameter in  $O(N \cdot 2^K)$ ,  $K = \max_i |pa(X_i)|$
  - ◆ Hinzufügen von Kanten: Familie der darstellbare Verteilungen wird grösser,  $I(G)$  wird kleiner
- Extremfälle: Graph ohne Kanten, (ungerichtet) vollständig verbundener Graph



$N$  Parameter

$$I(G) = \{(A \perp B | C) : A, B, C \text{ disj. Mengen von ZV}\}$$



$2^N - 1$  Parameter

$$I(G) = \emptyset$$

# Überblick

- Gerichtete Graphische Modelle: Bayessche Netze
- **Graphische Modelle im Maschinellen Lernen**
- Inferenz in Graphischen Modellen
- Ungerichtete Graphische Modelle: Markov Netze