Kapitel 1: Bayes-Netze

Bayes-Netze

Ausgangspunkt:

- Wahrscheinlichkeitsverteilungen über Welten sind geeignet, um unsicheres Wissen zu repräsentieren
- Eine naive Repräsentation als Liste von Wahrscheinlichkeiten ist
  - exponentiell groß, also mit vertretbarem Aufwand weder anzugeben noch zu speichern
  - auch für Experten wegen exotischer Sonderfälle nur schwer zu konstruieren
- Unabhängigkeit erlaubt kompakte Repräsentation und bringt Stuktur in die Menge der Ereignisse

Übersicht Kapitel 1

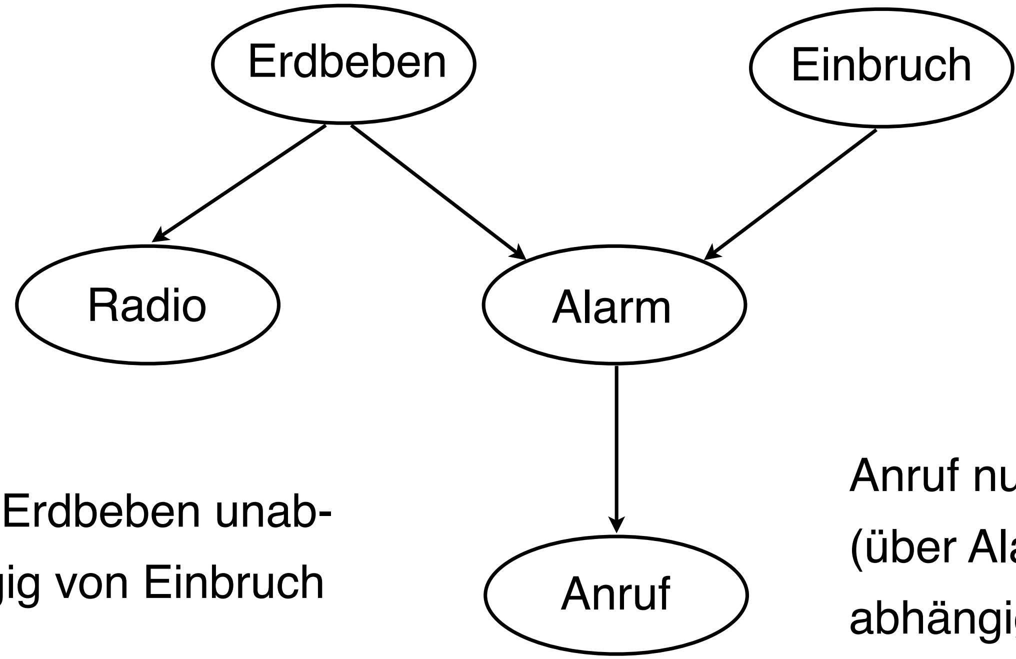
- Kapitel 1.1: Mehr zu Unabhängigkeit
- Kapitel 1.2: Bayes-Netze
- Kapitel 1.3: Graphoid Axiome
- Kapitel 1.4: d-Separation

Bayes-Netze

Kapitel 1.1: Mehr zu Unabhängigkeit

Bayes-Netze

Bayes-Netze verwenden Graph, um Unabhängigkeiten zu spezifizieren:



Z.B. Erdbeben unab-
hängig von Einbruch

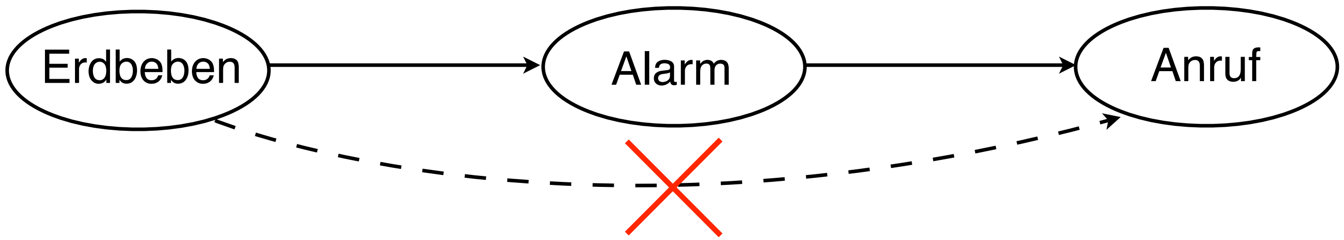
Anruf nur mittelbar
(über Alarm)
abhängig von Erdbeben

Basierend auf einer solchen Struktur kann man dann Verteilungen
in (meist) kompakter Weise beschreiben

Unabhängigkeit

Wir werden sehen: je weniger Pfeile im Unabhängigkeits-Graph, desto kompakter kann eine Verteilung für diesen Graph repräsentiert werden.

Mittelbare (Un)Abhängigkeiten wichtig zur Reduktion der Kantenzahl:



Formal beschreiben wir das mittels konditionaler Unabhängigkeit

Zentrale Beobachtung: Genau wie Wkten ist auch Unabhängigkeit von dynamischer Natur, kann von neuer Evidenz sowohl zerstört als auch hergestellt werden

Unabhängigkeit

Einbruch ist unabhängig von Erdbeben:

$\text{Pr}(\text{Erdbeben}) = .1$

Welt	Erdbeben	Einbruch	Alarm	$\text{Pr}(\cdot)$		$\text{Pr}(\cdot \mid \text{Erdbeben})$
ω_1	false	false	false	.7128		0
ω_2	false	false	true	.0072		0
ω_3	false	true	false	.0180	.1800	0
ω_4	false	true	true	.1620		0
ω_5	true	false	false	.0240		.2400
ω_6	true	false	true	.0560		.5600
ω_7	true	true	false	.0010	.0200	.0100
ω_8	true	true	true	.0190		.1900

$\text{Pr}(\text{Einbruch}) = .2$ $\text{Pr}(\text{Einbruch} \mid \text{Erdbeben}) = .2$

Unabhängigkeit

Nach Konditionierung mit Alarm ist das nicht mehr der Fall:

$$\Pr(\text{Einbruch}|\text{Alarm}) \approx .741$$

$$\Pr(\text{Einbruch}|\text{Alarm} \wedge \text{Erdbeben}) \approx .253$$

erst mit Alarm konditionieren, dann mit Erdbeben
(oder andersrum, was äquivalent ist)

Macht intuitiv Sinn:

- Einbruch und Erdbeben sind konkurrierende Ursachen für Alarm
- wenn wir lernen, dass eine davon eingetreten ist, nehmen wir an, dass die andere wohl nicht (zusätzlicher) Auslöser für den Alarm ist

Unabhängigkeit

Unabhängigkeit kann auch durch Konditionierung entstehen:

Angenommen wir haben zwei fehleranfällige Temperatursensoren, interessieren uns dafür, ob Temperatur normal oder extrem ist

Variablen: TNormal, S1Normal, S2Normal

Eingangs könnten wir haben:

$$\mathsf{Pr}(\mathsf{TNormal}) = .80$$

$$\mathsf{Pr}(\mathsf{S1Normal}) = .76$$

$$\mathsf{Pr}(\mathsf{S2Normal}) = .68$$

Mit Evidenz TNormal verschwindet diese Abhängigkeit:

$$\mathsf{Pr}(\mathsf{S2Normal}|\mathsf{TNormal}) = .80$$

$$\mathsf{Pr}(\mathsf{S2Normal}|\mathsf{TNormal} \wedge \mathsf{S1Normal}) = .80$$

Intuitiv sollte S2Normal abhängig sein von S1Normal:

$$\mathsf{Pr}(\mathsf{S2Normal}|\mathsf{S1Normal}) \approx .768$$

Konditionale Unabhängigkeit

Konditionale Unabhängigkeit von φ und ψ gegeben ϑ:
φ ist unabhängig von ψ nach Konditionierung mit ϑ

Definition konditional unabhängig

Seien φ, ψ, ϑ aussagenlogische Formeln. Wir nennen φ *unabhängig von* ψ *gegeben* ϑ wenn $\mathsf{Pr}(\varphi|\psi \wedge \vartheta) = \mathsf{Pr}(\varphi|\vartheta)$ oder $\mathsf{Pr}(\psi \wedge \vartheta) = 0$.

schreiben wir statt "$\mathsf{Pr}((\varphi|\psi)|\vartheta)$"

Äquivalent ist: $\mathsf{Pr}(\varphi \wedge \psi|\vartheta) = \mathsf{Pr}(\varphi|\vartheta) \cdot \mathsf{Pr}(\psi|\vartheta)$

Dies zeigt auch folgende Symmetrie:
φ ist unabhängig von ψ gegeben ϑ gdw. ψ ist unabhängig von φ gegeben ϑ.

Unabhängigkeit von Variablenmengen

Wir wollen in der Lage sein, zu beschreiben, dass es innerhalb einer Menge von Variablen keinerlei (unkonditionale) Abhängigkeiten gibt

Bereits gesehen: paarweise Unabhängigkeit ist nicht ausreichend

Definition *I*(X,Z,Y)

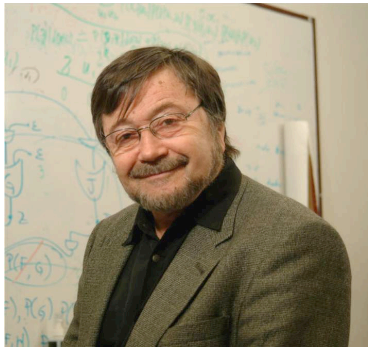
Seien X, Y, Z disjunkte Mengen von Variablen. Dann ist X *unabhängig von* Y *gegeben* Z, geschrieben $I_{\mathsf{Pr}}(X, Z, Y)$ wenn folgendes gilt:

> für alle $\varphi_{X'} \in \mathsf{Inst}(X)$, $\varphi_{Y'} \in \mathsf{Inst}(Y)$, $\varphi_{Z'} \in \mathsf{Inst}(Z)$:
> $\varphi_{X'}$ ist unabhängig von $\varphi_{Y'}$ gegeben $\varphi_{Z'}$.

Wenn X (oder Y oder Z) nur ein Element hat, lassen wir Mengenklammern weg, z.B. $I_{\mathsf{Pr}}(x, y, \{z_1, z_2\})$ statt $I_{\mathsf{Pr}}(\{x\}, \{y\}, \{z_1, z_2\})$.

Bayes-Netze

Kapitel 1.2: Bayes-Netze



Bayes-Netze

Ein Bayes-Netz besteht aus

- gerichtetem azyklischen Graph (DAG), der Unabhängigkeiten beschreibt
- Annotation dieses Graphen mit Wahrscheinlichkeiten

deren Kombination genau eine Verteilung definiert.

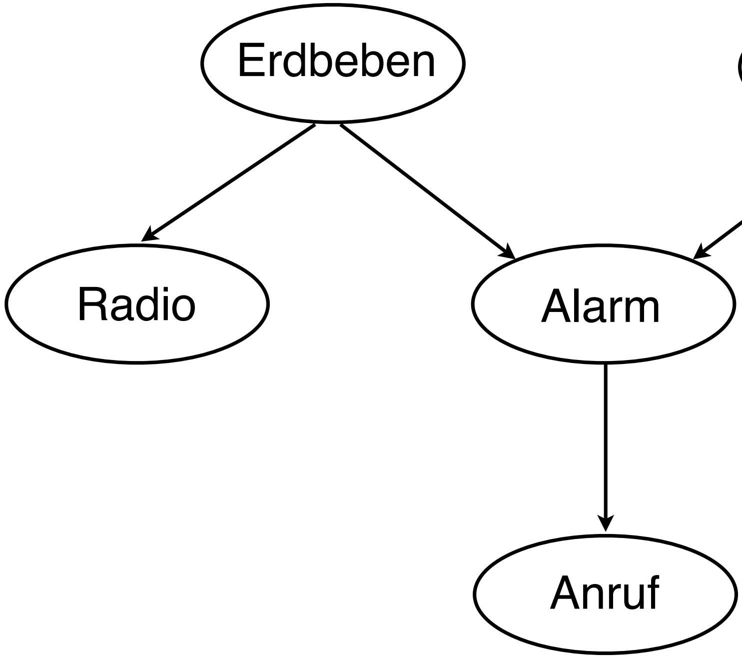
Notation für Graphen: für jede Variable x ist

$\mathsf{Parents}(x)$ die Menge der Knoten y mit direkter Kante von y nach x

$\mathsf{Descendants}(x)$ die Menge der Knoten y mit einer Kantenfolge beliebiger Länge ≥ 1 von x nach y

$\mathsf{Non\text{-}Descendants}(x)$ sind alle Variablen außer x, $\mathsf{Parents}(x)$ und $\mathsf{Descendants}(x)$

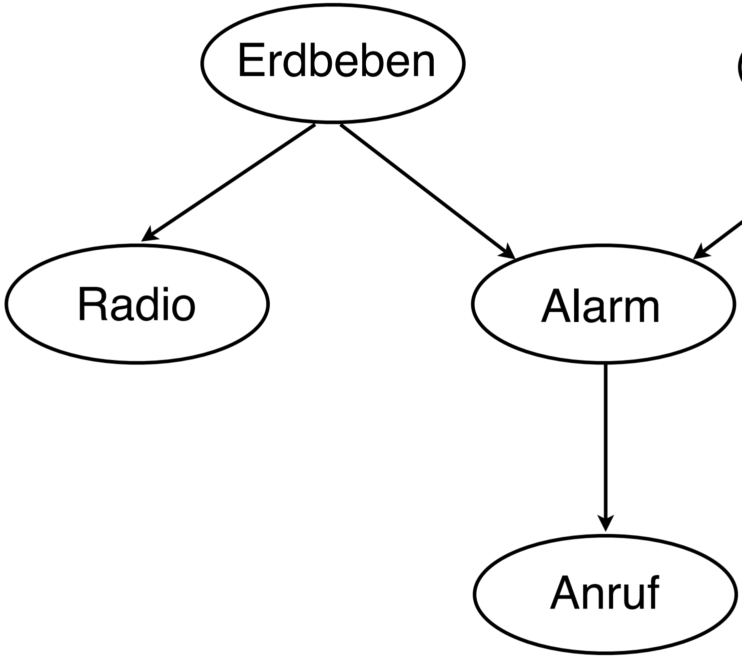
Bayes-Netze



$\text{Parents}(\text{Erdbeben}) = \emptyset$ $\text{Parents}(\text{Alarm}) = \{\text{Erdbeben}, \text{Einbruch}\}$

$\text{Parents}(\text{Anruf}) = \{\text{Alarm}\}$

Bayes-Netze



$$\text{Non-Descendants}(\mathsf{Erdbeben}) = \{\mathsf{Einbruch}\}$$

$$\text{Non-Descendants}(\mathsf{Alarm}) = \{\mathsf{Erdbeben}, \mathsf{Einbruch}, \mathsf{Radio}\}$$

$$\text{Non-Descendants}(\mathsf{Anruf}) = \{\mathsf{Erdbeben}, \mathsf{Einbruch}, \mathsf{Radio}, \mathsf{Alarm}\}$$

Bayes-Netze

Ein DAG repräsentiert die folgenden Unabhängigkeiten:

$$I(x, \mathsf{Parents}(x), \mathsf{Non\text{-}Descendants}(x)) \qquad \text{für jeden Knoten } x$$

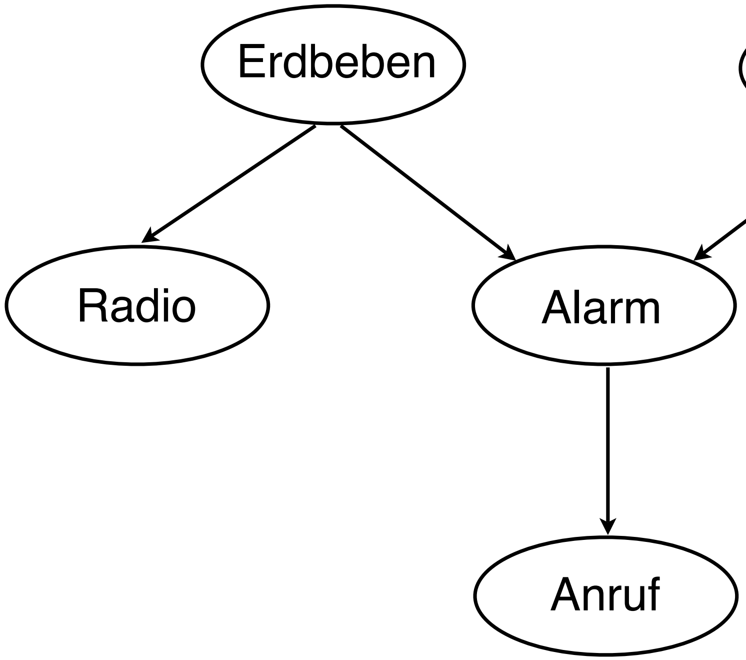
In Worten:

jedes Ereignis x ist unabhängig von $\mathsf{Non\text{-}Descendants}(x)$ gegeben $\mathsf{Parents}(x)$.

Idee dahinter:

- Der Übersichtlichkeit halber lesen wir Graph von oben nach unten
  blenden alle Descendants erstmal aus; darum Non-Descendants
- Abhängigkeiten von anderen Knoten kann es nur nur mittelbar
  über die Parents geben
- aus den Wkten von $\mathsf{Parents}(x)$ ergibt sich also die Wkt von x,
  die Wkt anderer Knoten spielt dann keine Rolle mehr

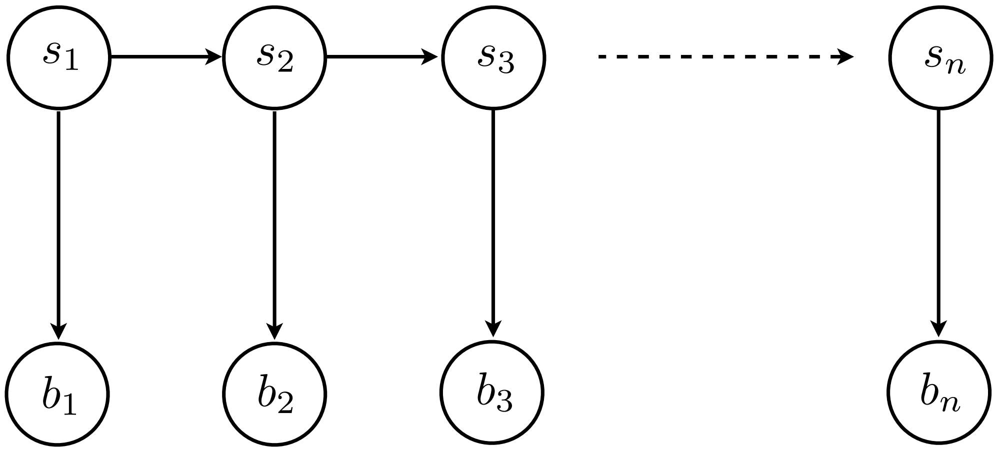
Bayes-Netze



Für einen DAG G bezeichnen wir die Menge aller Statements $I(x, \text{Parents}(x), \text{Non-Descendants}(x))$ mit $\text{Unabh}(G)$.

Bayes-Netze

Folgender DAG heisst Hidden Markov Model (HMM):



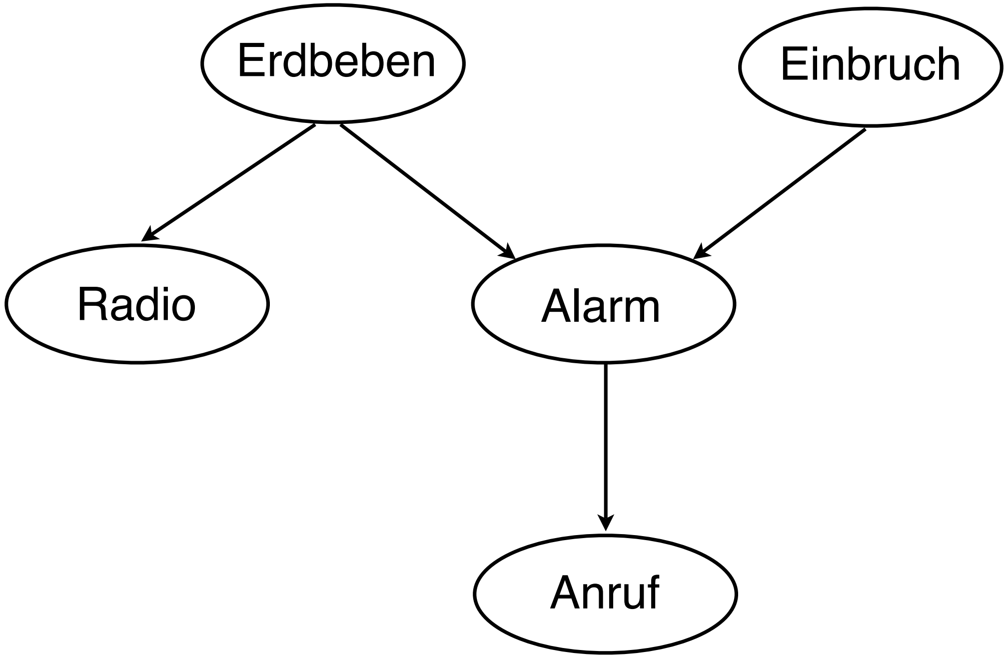
Das HMM repräsentiert die Evolution eines Systems von Zeitpunkt 1 bis n

Dabei bezeichnet s_i den tatsächlichen Zustand des Systems
und b_i der Wert eines Sensors, der das System beobachtet
(alle Variablen mehrwertig)

Für jedes s_i ergibt sich $I(s_i, s_{i-1}, \{s_1, \ldots, s_{i-2}, b_1, \ldots, b_{i-1}\})$.

Bayes-Netze

Jede Variable x annotiert mit konditionaler Wahrscheinlichkeitstabelle (KWT): für jede Belegung von $\mathsf{Parents}(x)$ eine Verteilung über die Werte von x.



Zusammen mit den Unabhängigkeiten in $\mathsf{Unabh}(G)$ definieren die KWTen eine eindeutige Verteilung Pr über allen Variablen in G

Bayes-Netze

Die Größe einer KWT ist natürlich im Prinzip immernoch exponentiell

Allerdings nur in der Anzahl der Parents (meist wenige)
statt in der Anzahl aller Variablen (meist viele)

Im konkreten Fall des Einbruch-Erdbeben-Alarm-DAGs:

- alle KWTs zusammengenommen enthalten 10 Wkten
- es gibt 5 Variablen, also 32 Belegungen und bei naiver Repräsentation sind demnach 32 Wkten enzugeben

Im folgenden: Formale Definition der Syntax und Semantik

Bayes-Netze

Definition Bayes-Netz

Ein *Bayes-Netz (BN)* ist ein Paar $N = (G, \Theta)$ wobei

- G die *Struktur* von N ist:
  gerichteter azyklischer Graph, dessen Knoten wir Variablen nennen
- Θ die *Parametrisierung* von N ist:
  Eine konditionale Wahrscheinlichkeitstabelle für jede Variable

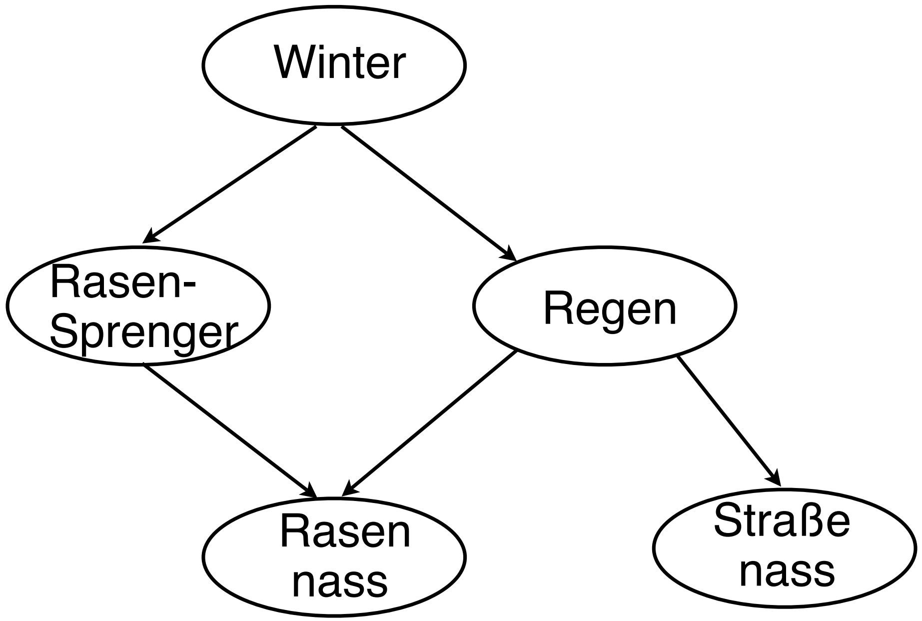
Wir bezeichnen mit

- $\Theta_{x|P}$ die KWT für die Variable x mit $\mathsf{Parents}(x) = P$
- $\theta_{x=v|\omega}$ die Wkt für $x = v$ in der Zeile von $\Theta_{x|P}$ für Belegung ω

  (z.B. $\theta_{\mathsf{Alarm}=\mathsf{an}|\mathsf{Erdbeben}\wedge\neg\mathsf{Einbruch}}$)

Bayes-Netze

Ein weiteres Beispiel:

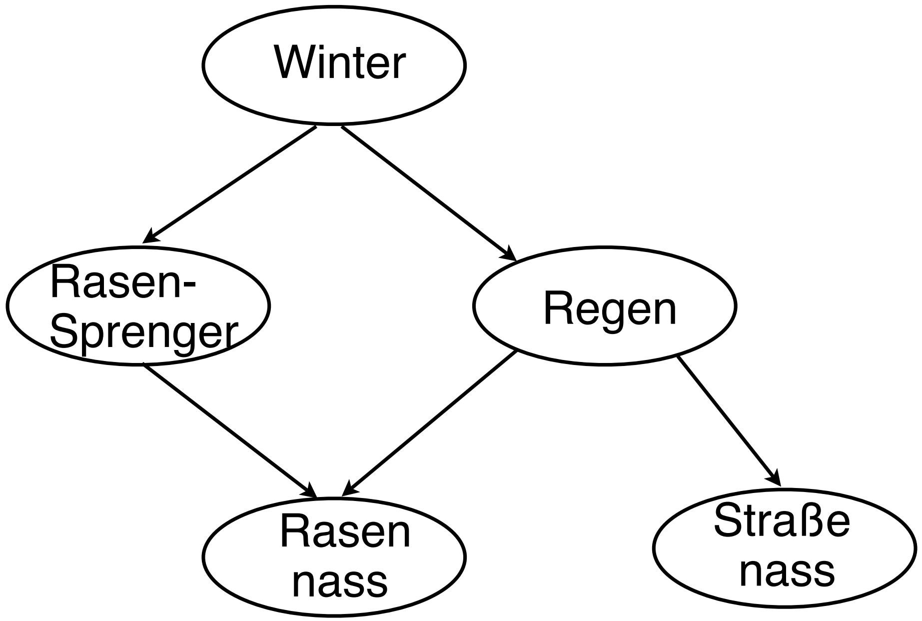


$\Theta_{\mathsf{Winter}|\emptyset}$ ist

Winter $=$ true	Winter $=$ false
.6	.4

Bayes-Netze

Ein weiteres Beispiel:

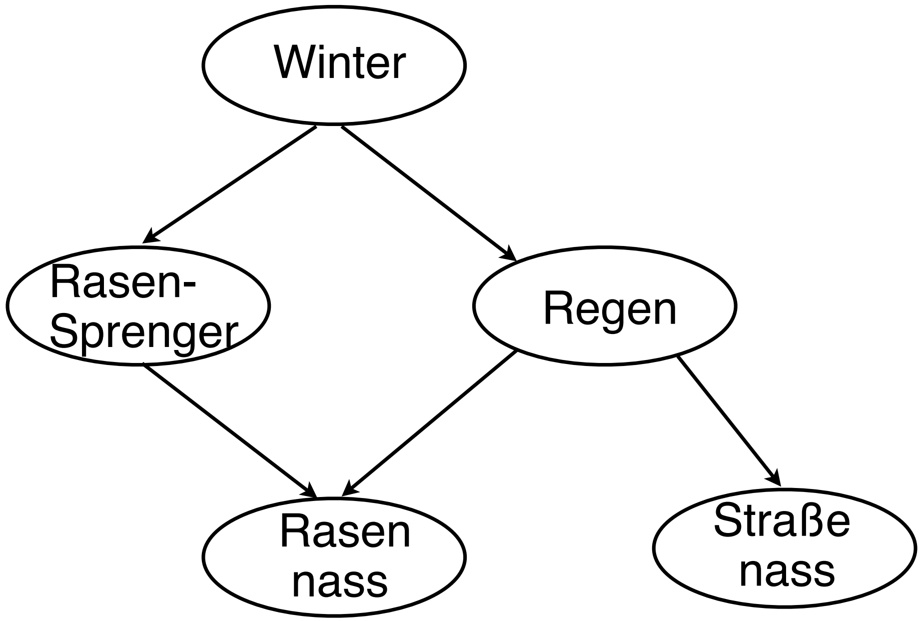


$\Theta_{\text{Rasensprenger}|\text{Winter}}$ ist

Winter	Rasensprenger $=$ true	Rasensprenger $=$ false
t	.2	.8
f	.75	.25

Bayes-Netze

Ein weiteres Beispiel:

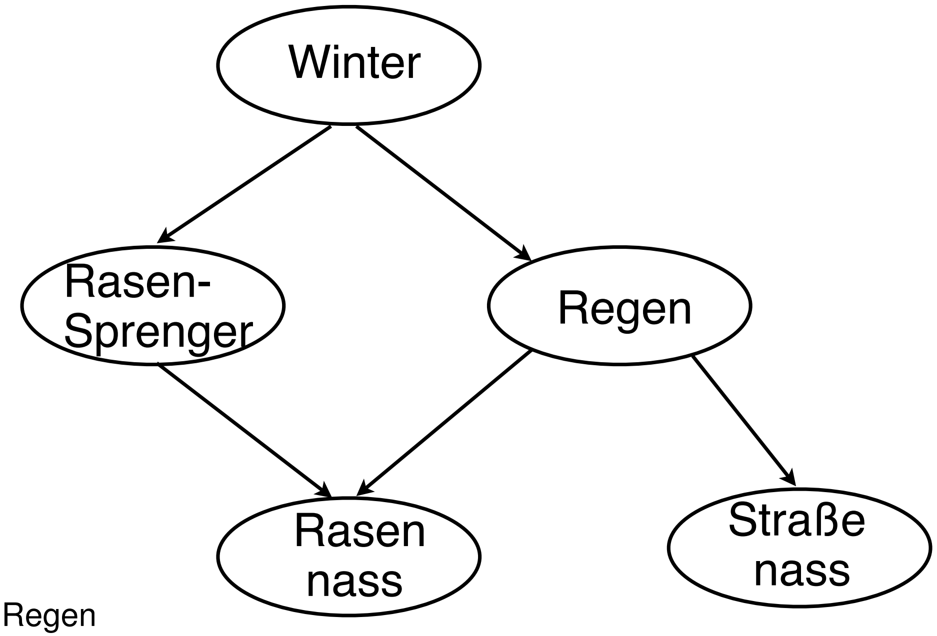


$\Theta_{\text{Regen}|\text{Winter}}$ ist

Winter	Regen	¬Regen
t	.8	.2
f	.1	.9

Bayes-Netze

Ein weiteres Beispiel:



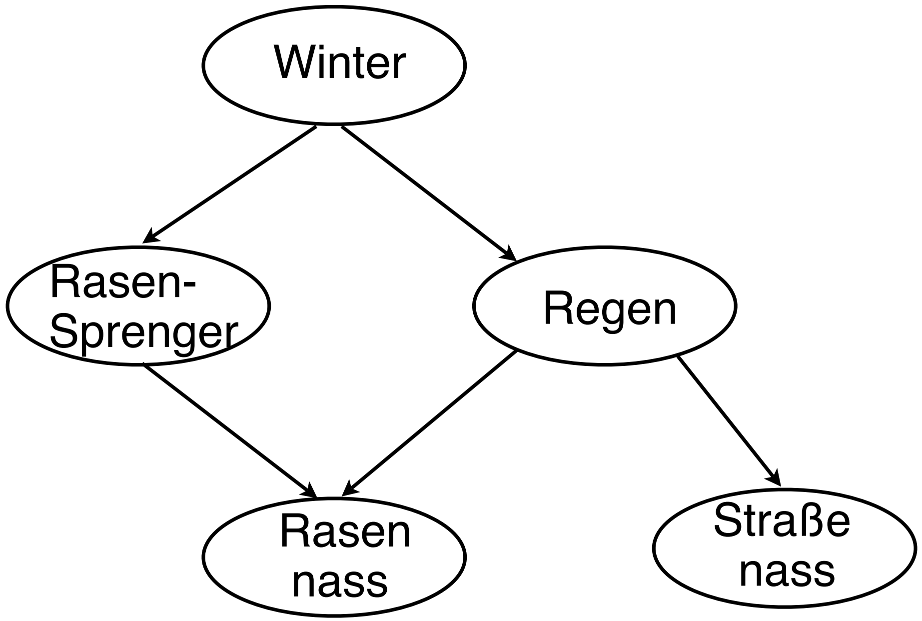
$\Theta_{\text{Rasen nass}|\text{Rasensprenger,Regen}}$

ist

Rasensprenger	Regen	Rasennass	¬Rasennass
t	t	.95	.05
t	f	.9	.1
f	t	.8	.2
f	f	0	1

Bayes-Netze

Ein weiteres Beispiel:

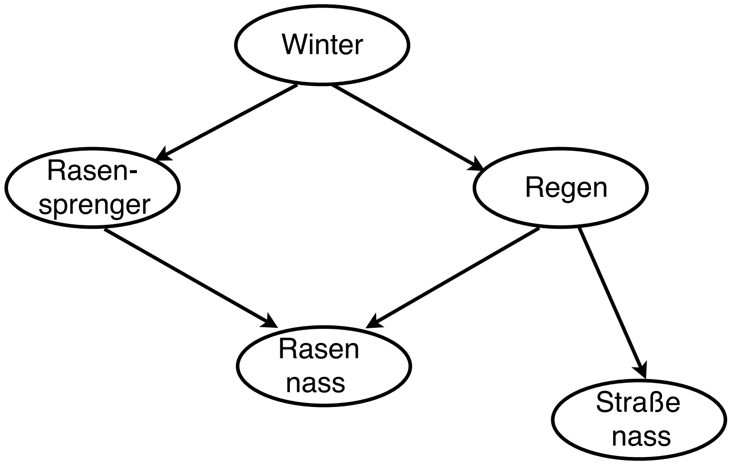


$\Theta_{\text{Strasse nass}|\text{Regen}}$ ist

Regen	Strassenass	¬Strassenass
t	.7	.3
f	0	1

Bayes-Netze: Semantik

Die durch ein BN repräsentierte Verteilung illustrieren wir zunächst anhand eines Beispiels



Betrachte Belegung

$$
\begin{aligned}
\omega(\mathsf{W}) &= \text{true}\\
\omega(\mathsf{RS}) &= \text{false}\\
\omega(\mathsf{R}) &= \text{true}\\
\omega(\mathsf{Rn}) &= \text{true}\\
\omega(\mathsf{Sn}) &= \text{false}
\end{aligned}
$$

$\Theta_{\mathsf{W}|\emptyset}$

W	¬W
.6	.4

$\Pr(\mathsf{W}) = .6$

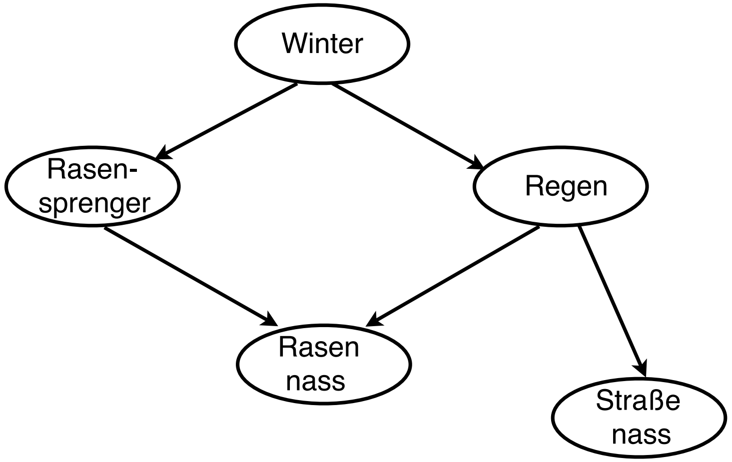
$\Theta_{\mathsf{RS}|\mathsf{W}}$

W	RS	¬RS
t	.2	.8
f	.75	.25

$\Pr(\mathsf{W} \wedge \neg\mathsf{RS}) = .6 \cdot .8 = .48$

Bayes-Netze: Semantik

Die durch ein BN repräsentierte Verteilung illustrieren wir zunächst anhand eines Beispiels



Betrachte Belegung

$$
\begin{aligned}
\omega(\mathsf{W}) &= \mathsf{true}\\
\omega(\mathsf{RS}) &= \mathsf{false}\\
\omega(\mathsf{R}) &= \mathsf{true}\\
\omega(\mathsf{Rn}) &= \mathsf{true}\\
\omega(\mathsf{Sn}) &= \mathsf{false}
\end{aligned}
$$

$$\mathsf{Pr}(\mathsf{W} \wedge \neg\mathsf{RS}) = .48$$

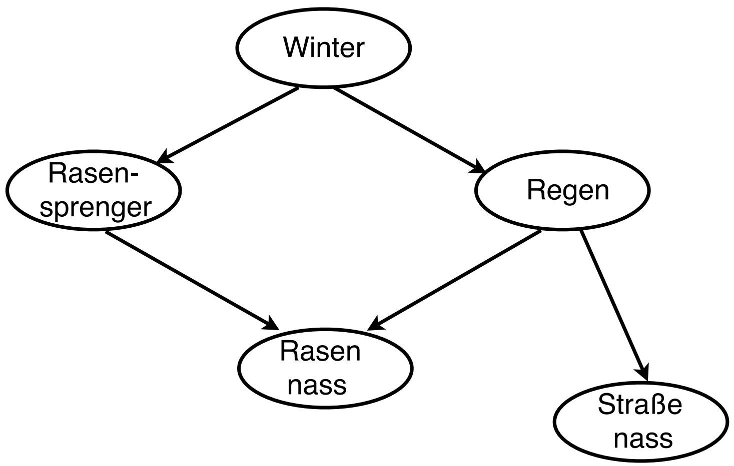
$\Theta_{\mathsf{R}|\mathsf{W}}$

W	R	¬R
t	.8	.2
f	.1	.9

$$\mathsf{Pr}(\mathsf{W} \wedge \neg\mathsf{RS} \wedge \mathsf{R}) = .48 \cdot .8 = .384$$

Bayes-Netze: Semantik

Die durch ein BN repräsentierte Verteilung illustrieren wir zunächst anhand eines Beispiels



Betrachte Belegung

$$
\begin{aligned}
\omega(\mathsf{W}) &= \mathsf{true}\\
\omega(\mathsf{RS}) &= \mathsf{false}\\
\omega(\mathsf{R}) &= \mathsf{true}\\
\omega(\mathsf{Rn}) &= \mathsf{true}\\
\omega(\mathsf{Sn}) &= \mathsf{false}
\end{aligned}
$$

$\Theta_{\mathsf{Rn}|\mathsf{RS},\mathsf{R}}$

RS	R	Rn	¬Rn
t	t	.95	.05
t	f	.9	.1
f	t	.8	.2
f	f	0	1

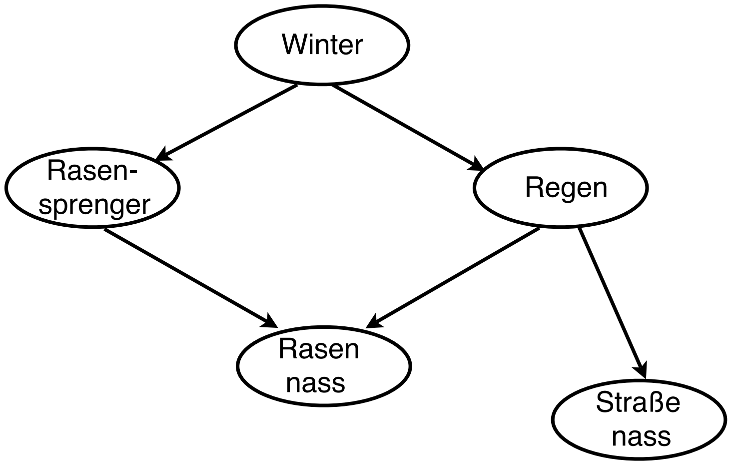
$$\mathsf{Pr}(\mathsf{W} \wedge \neg\mathsf{RS} \wedge \mathsf{R}) = .384$$

$$\mathsf{Pr}(\mathsf{W} \wedge \neg\mathsf{RS} \wedge \mathsf{R} \wedge \mathsf{Rn})$$

$$= .384 \cdot .8 = .3072$$

Bayes-Netze: Semantik

Die durch ein BN repräsentierte Verteilung illustrieren wir zunächst anhand eines Beispiels



Betrachte Belegung

$$
\begin{aligned}
\omega(\mathsf{W}) &= \mathsf{true}\\
\omega(\mathsf{RS}) &= \mathsf{false}\\
\omega(\mathsf{R}) &= \mathsf{true}\\
\omega(\mathsf{Rn}) &= \mathsf{true}\\
\omega(\mathsf{Sn}) &= \mathsf{false}
\end{aligned}
$$

$\Theta_{\mathsf{Sn}|\mathsf{R}}$

R	Sn	¬Sn
t	.7	.3
f	0	1

$$\mathsf{Pr}(\mathsf{W} \wedge \neg\mathsf{RS} \wedge \mathsf{R} \wedge \mathsf{Rn}) = .3072$$

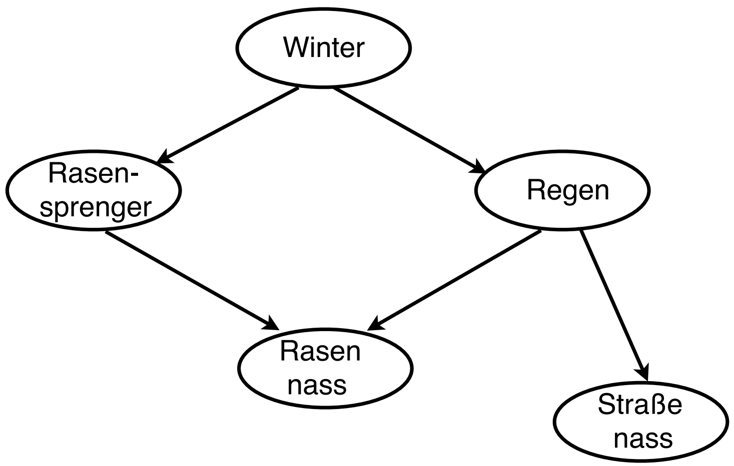
$$\mathsf{Pr}(\mathsf{W} \wedge \neg\mathsf{RS} \wedge \mathsf{R} \wedge \mathsf{Rn} \wedge \mathsf{Sn})$$

$$= .3072 \cdot .7 = .21504$$

$$= \mathsf{Pr}(\omega)\ !$$

Bayes-Netze: Semantik

Die durch ein BN repräsentierte Verteilung illustrieren wir zunächst anhand eines Beispiels



Betrachte Belegung

$$\begin{aligned} \omega(\mathsf{W}) &= \mathsf{true} \\ \omega(\mathsf{RS}) &= \mathsf{false} \\ \omega(\mathsf{R}) &= \mathsf{true} \\ \omega(\mathsf{Rn}) &= \mathsf{true} \\ \omega(\mathsf{Sn}) &= \mathsf{false} \end{aligned}$$

Zusammengefasst also

$$\mathsf{Pr}(\omega) = \theta_{\mathsf{W}|\cdot} \cdot \theta_{\neg\mathsf{RS}|\mathsf{W}} \cdot \theta_{\mathsf{R}|\mathsf{W}} \cdot \theta_{\mathsf{Rn}|\mathsf{RS}\wedge\mathsf{R}} \cdot \theta_{\neg\mathsf{Sn}|\mathsf{R}}$$

Dies entspricht im wesentlichen der sog. Kettenregel für Wkten:

$$\mathsf{Pr}(\varphi_1 \wedge \cdots \wedge \varphi_n) = \mathsf{Pr}(\varphi_1|\varphi_2 \wedge \cdots \wedge \varphi_n)\mathsf{Pr}(\varphi_2|\varphi_3 \wedge \cdots \wedge \varphi_n) \cdots \mathsf{Pr}(\alpha_n).$$

Bayes-Netze: Semantik

Formal definiert man die von einem BN repräsentierte Belegung völlig analog zu vorigem Beispiel

Definition *Verteilung eines Bayes-Netzes*

BN $N = (G, \Theta)$ mit Variablen $x_1, \ldots, x_n$ repräsentiert Verteilung Pr_N:
für jede Belegung $\omega = \{x_i \mapsto v_i \mid 1 \leq i \leq n\}$,

$$\mathsf{Pr}_N(\omega) = \prod_{i=1..n} \theta_{x_i = v_i \,|\, \omega/\mathsf{Parents}(x_i)}$$

mit ω/X Einschränkung von ω auf die Variablen aus der Menge X.

Vergleiche mit

$$\mathsf{Pr}(\omega) = \theta_{\mathsf{W}|\cdot} \cdot \theta_{\neg\mathsf{RS}|\mathsf{W}} \cdot \theta_{\mathsf{R}|\mathsf{W}} \cdot \theta_{\mathsf{Rn}|\mathsf{RS}\wedge\mathsf{R}} \cdot \theta_{\neg\mathsf{Sn}|\mathsf{R}}$$

Das BN N stellt also eine Faktorisierung der Verteilung Pr_N dar.

Bayes-Netze: Semantik

Die Belegung Pr_N verhält sich wie erwartet (Beweis als Aufgabe):

Lemma

(V1) Pr_N erfüllt alle Unabhängigkeiten in $\mathsf{Markov}(G)$

(V2) Pr_N erfüllt Θ in folgendem Sinne:

jedes $\theta_{x=v|\omega}$ ist die konditionale Wkt für $x = v$ gegeben φ_ω, also:

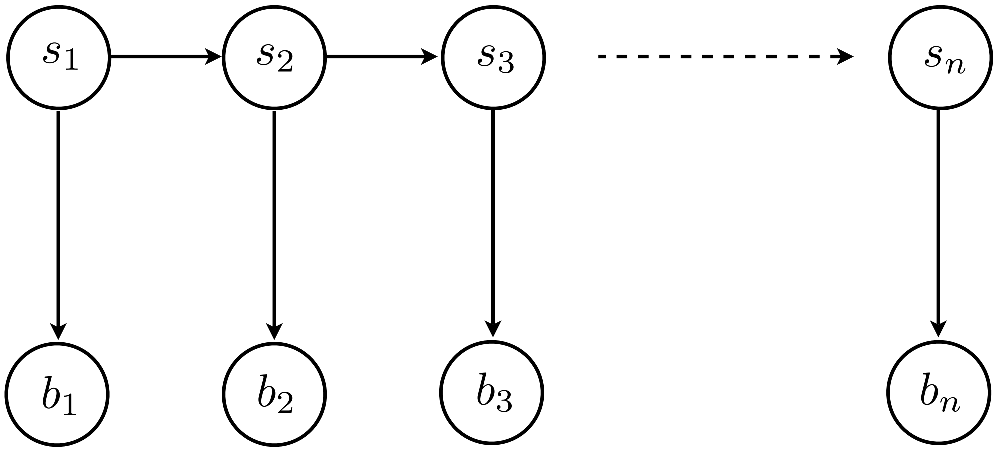
$$\mathsf{Pr}(x = v|\varphi_\omega) = \theta_{x=v|\omega}$$

Wir werden später sogar zeigen:

Pr_N ist die einzige Belegung, die (V1) und (V2) erfüllt!

Bayes-Netze: Semantik

KWTs im Hidden Markov Model mit m Zuständen und m Sensorwerten

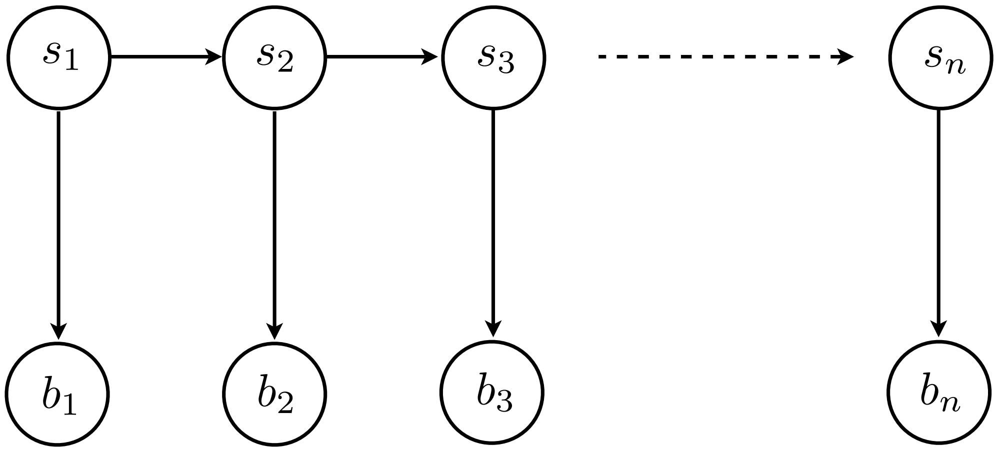


Die KWT für s_1 besteht aus einer einzelnen Verteilung:

$s_1 = 1$	$s_1 = 2$	$\cdots$	$s_1 = m$
.1	.28	$\cdots$	.4

Bayes-Netze: Semantik

KWTs im Hidden Markov Model mit m Zuständen und m Sensorwerten

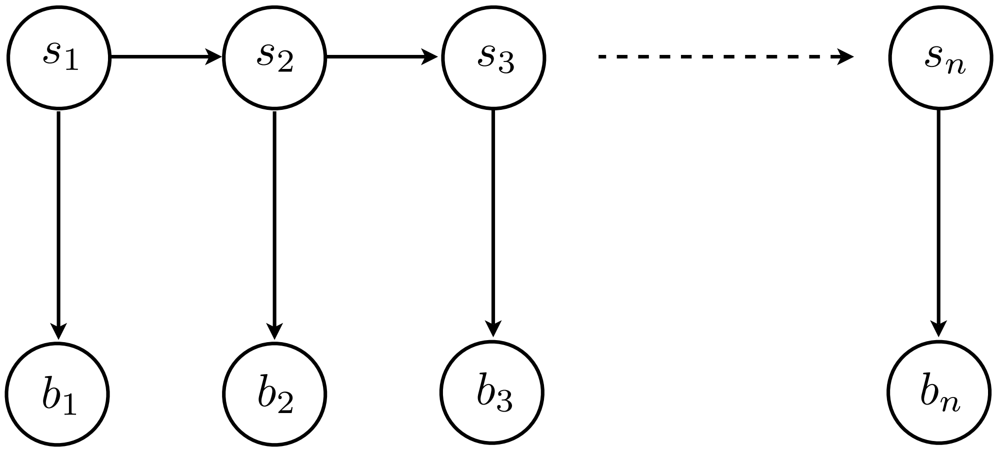


Die KWTen für $s_2, \ldots, s_n$ haben m^2 Einträge und sind alle identisch

Wert s_{i-1}	$s_i = 1$	$\cdots$	$s_i = m$
1	.21	$\cdots$	.05
$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	0	$\cdots$	1

Bayes-Netze: Semantik

KWTs im Hidden Markov Model mit m Zuständen und m Sensorwerten



Die KWTen für $b_1, \ldots, b_n$ haben m^2 Einträge und sind alle identisch

Wert s_i	$b_i = 1$	$\cdots$	$b_i = m$
1	.9	$\cdots$	.05
$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	.02	$\cdots$	.91

Bayes-Netze

Kapitel 1.3: Graphoid Axiome

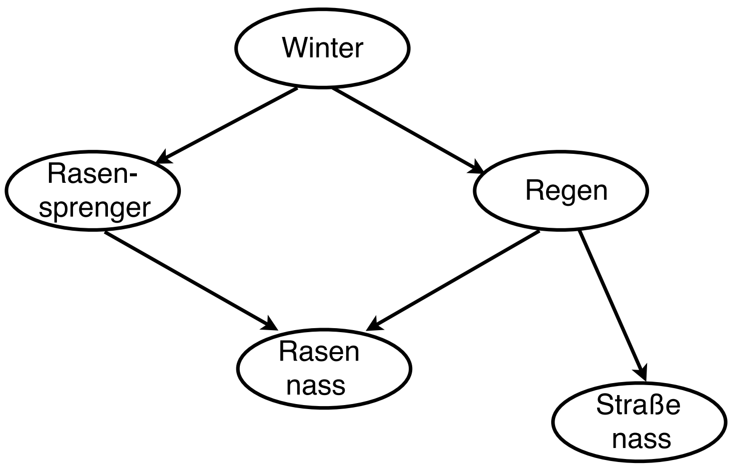
Unabhängigkeit

Unabhängigkeit zental in Bayes-Netzen / Faktorisierungen von Verteilungen

Wir wissen bereits:

Für $N = (G, \Theta)$ erfüllt Pr_N alle Unabhängigkeiten in $\mathsf{Markov}(G)$.

$\mathsf{Markov}(G)$ sind jedoch keineswegs die einzigen Unabhängigkeiten in Pr_N:



In $\mathsf{Markov}(G)$:

$I(\text{Straßenass}, \text{Regen}, \{\text{Winter}, \text{Rasensprenger}, \text{Rasennass}\})$

dann aber auch

$I(\text{Straßenass}, \text{Regen}, \text{Winter})$

und — nicht in $\mathsf{Markov}(G)$

$I(\text{Winter}, \text{Regen}, \text{Straßenass})$

Graphoid Axiome

Frage also: Welche Unabhängigkeiten gelten (noch) in Pr_N?

Die Graphoid Axiome beschreiben, wie sich aus einer Menge konditionaler Unabhängigkeiten weitere konditionale Unabhängigkeiten ergeben

Axiomschemata I: Symmetrie

Lemma

Für alle Verteilungen Pr gilt: wenn $I_{\text{Pr}}(X, Z, Y)$, dann $I_{\text{Pr}}(Y, Z, X)$.

Ergibt sich direkt aus alternativer Def. von Unabhängigkeit:

$$\text{Pr}(\varphi \wedge \psi | \vartheta) = \text{Pr}(\varphi | \vartheta) \cdot \text{Pr}(\psi | \vartheta)$$

Graphoid Axiome

Axiomschema II: Dekomposition

Lemma

Für alle Pr gilt: wenn $I_{\mathsf{Pr}}(X, Z, Y)$ und $Y' \subseteq Y$, dann $I_{\mathsf{Pr}}(X, Z, Y')$

Beweis: Übung

Lemma

Pr_N ist die einzige Belegung, die folgende Bedingungen erfüllt:

(V1) Pr_N erfüllt alle Unabhängigkeiten in $\mathsf{Markov}(G)$

(V2) Pr_N erfüllt Θ in folgendem Sinne:

jedes $\theta_{x=v|\omega}$ ist die konditionale Wkt für $x = v$ gegeben φ_ω, also:

$$\mathsf{Pr}(x = v|\varphi_\omega) = \theta_{x=v|\omega}$$

Graphoid Axiome

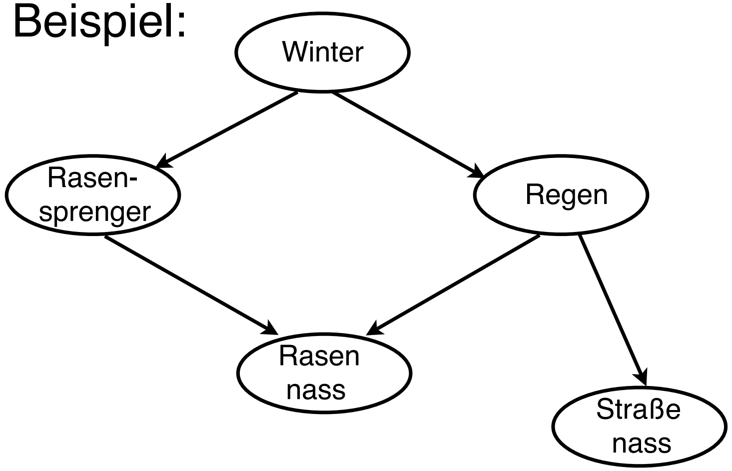
Axiomschema III: Schwache Vereinigung

Lemma

Für alle Pr gilt: wenn $I_{\mathsf{Pr}}(X, Z, Y \cup W)$, dann $I_{\mathsf{Pr}}(X, Z \cup Y, W)$

Intuitiv:
wenn $Y \cup W$ nicht relevant für X ist und wir die Werte der für Y "lernen", dann wird W dadurch nicht relevant

Beispiel:



In Markov(G):

$I($Straßenass, Regen, $\{$Winter, Rasensprenger, Rasennass$\})$

mit schwacher Vereinigung auch

$I($Straßenass, $\{$Regen, Winter$\}$, $\{$Rasensprenger, Rasennass$\})$

Graphoid Axiome

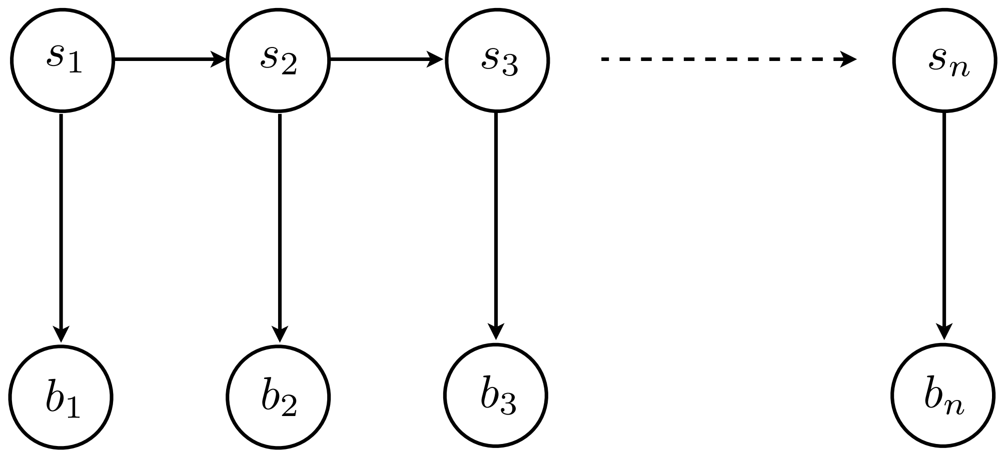
Axiomschema IV: Kontraktion

Lemma

Für alle Pr gilt: wenn $I_{\mathsf{Pr}}(X, Z, Y)$ und $I_{\mathsf{Pr}}(X, Z \cup Y, W)$, dann $I_{\mathsf{Pr}}(X, Z, Y \cup W)$

Intuitiv:
wenn wir die Werte der irrelevanten Variablen Y lernen und W danach irrelevant ist, dann war vorher bereits $Y \cup W$ irrelevant

Beispiel:



Graphoid Axiome

Axiomschema V: Schnitt

Lemma

Für alle positiven Pr (d.h. $\mathsf{Pr}(\omega) > 0$ für alle ω) gilt:
wenn $I_{\mathsf{Pr}}(X, Z \cup W, Y)$ und $I_{\mathsf{Pr}}(X, Z \cup Y, W)$, dann $I_{\mathsf{Pr}}(X, Z, Y \cup W)$

Intuitiv:
wenn Y nach lernen von W irrelevant ist und umgekehrt, dann $Y \cup W$ irrelevant

Positive Verteilungen sind

- nicht adäquat für streng logische Variablenzusammenhänge
  wie $\quad x \to y, \quad x_1 \wedge x_2 \to \neg y, \quad x_1 \vee x_2$
- für typische BN-Anwendungen aber durchaus realistisch

Das Schnitt Axiom ist in nicht-positiven Verteilungen nicht erfüllt

Graphoid Axiome

Liste der Graphoid Axiome:

I	Symmetrie
II	Dekomposition
III	Schwache Vereinigung
IV	Kontraktion
(V)	Schnitt - nur für positive Verteilungen

Manchmal wird noch das *Trivialitätsaxiom* $I_{\mathsf{Pr}}(X, Z, \emptyset)$ hinzugenommen.

Die Graphoid Axiome sind nicht vollständig im folgenden Sinn:

> Wenn jede Verteilung Pr, die Menge M von kond. Unabhängigkeiten erfüllt, auch $I(X, Z, Y)$ erfüllt, dann lässt sich $I(X, Z, Y)$ aus M mittels der Graphoid Axiome herleiten.

Man kann zeigen, dass keine endliche Axiomatisierung existiert mit Axiomen der Form $I(X_1, Z_1, Y_2) \wedge \cdots \wedge I(X_n, Z_n, Y_n) \Longrightarrow I(X, Z, Y)$

Kapitel 1.4: d-Separation

d-Separation

Ziel:

Unabhängigkeiten in einem BN in grafischer Weise charakterisieren und effiziente Algorithmen für folgendes Problem finden:

> Gegeben BN N und $I(X, Z, Y)$,
> entscheide ob "$I(X, Z, Y)$ Unabhängigkeit in N ist".

werden wir später präzise machen

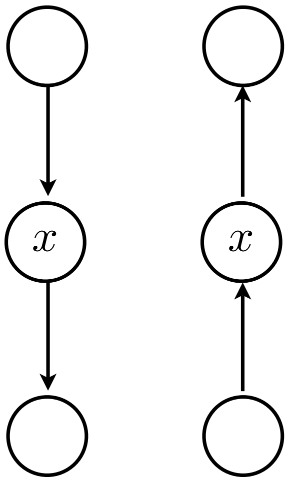
Grundidee:

$I(X, Z, Y)$ gilt in N wenn alle Pfade zwischen $x \in X$ und $y \in Y$ in geeigneter Weise durch Knoten aus Z "unterbrochen" sind.

d-Separation

Die Knoten auf dem Pfad stellt man sich am besten als Ventile vor

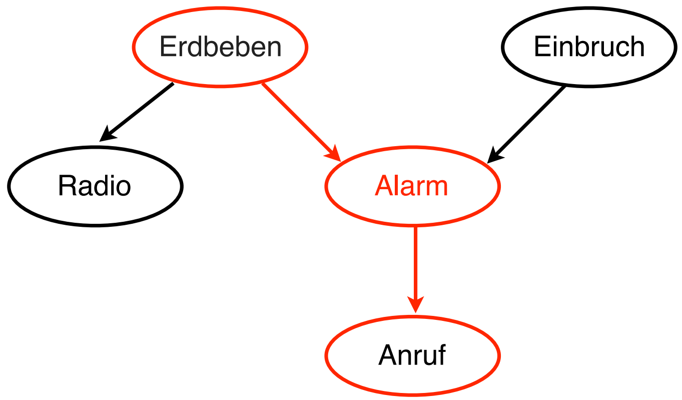
Es gibt drei Arten von Ventilen:



Sequentielles Ventil

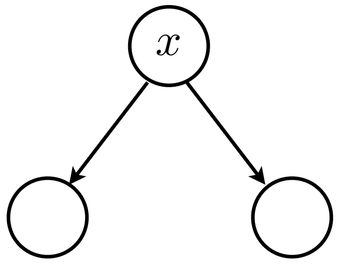
Geschlossen wenn

$x \in Z$



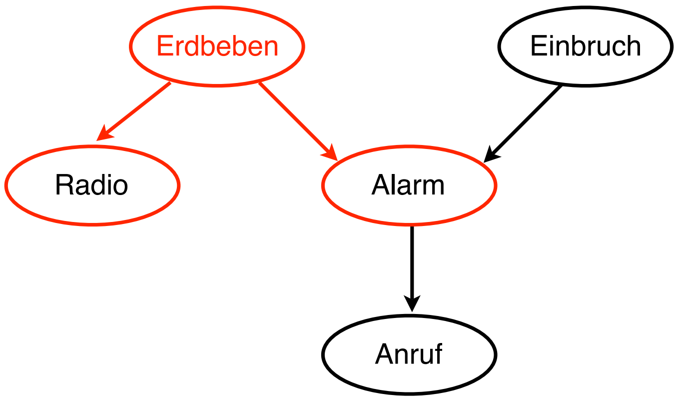
Erdbeben und Anruf unabhängig gdw Wert von Alarm bekannt

d-Separation



Divergentes Ventil

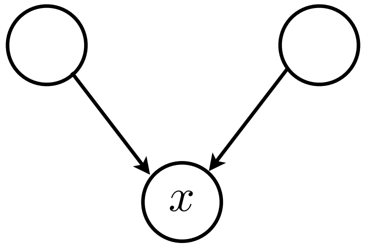
Geschlossen wenn $x \in Z$



1. Wert von Erdbeben bekannt
   $\Rightarrow$ Radio und Alarm unabhängig

2. Wert von Erdbeben unbekannt
   $\Rightarrow$ z.B. Alarm erhöht Wkt einer Radionachricht über Erdbeben

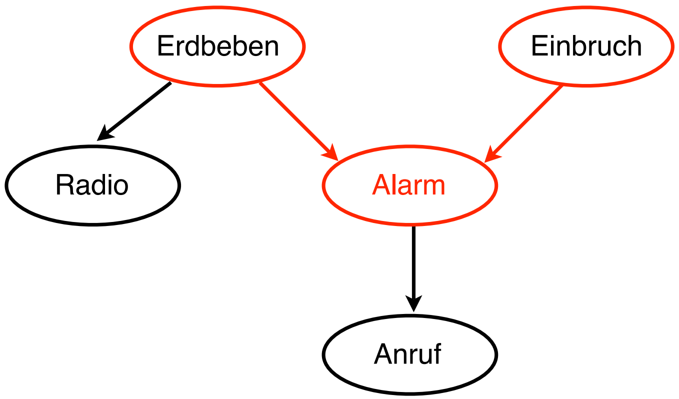
d-Separation



Konvergentes Ventil

Geschlossen wenn $x \notin Z$

und Descendants$(x) \cap Z = \emptyset$

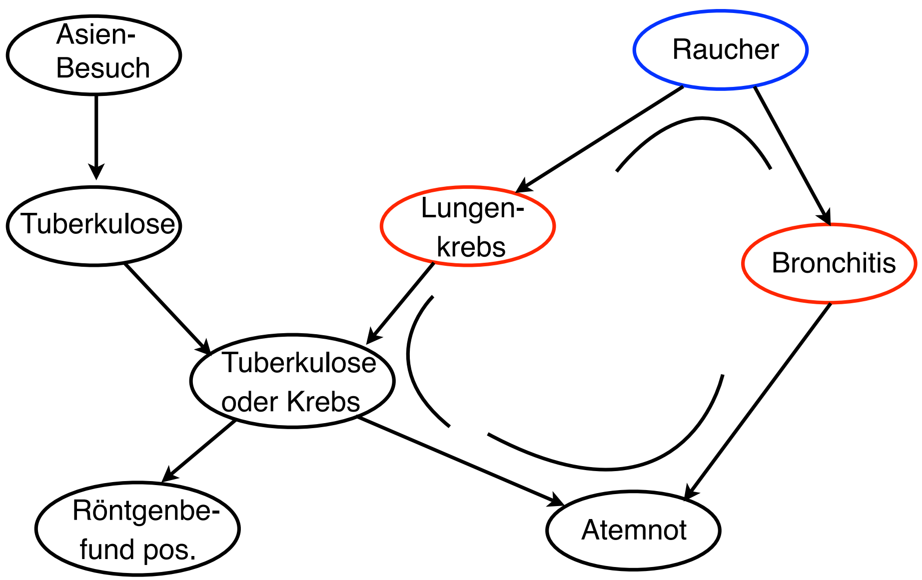


1. Wert von Alarm und Anruf unbekannt
   $\Rightarrow$ Erdbeben und Einbruch unabhängig
2. Wert von z.B. Alarm bekannt
   $\Rightarrow$ Erdbeben verringert Wkt von Einbruch

d-Separation

Definition d-separiert

Seien X, Y, Z disjunkte Knotenmengen in DAG G. X und Y sind *d-separiert durch* Z, geschrieben $\mathrm{dsep}_G(X, Z, Y)$, wenn auf jedem Pfad von $x \in X$ nach $y \in Y$ ein durch Z geschlossenes Ventil liegt. Der Pfad heißt dann Z*-blockiert*.



$\mathrm{dsep}_G($Bronchitis, Raucher, Lungenkrebs$)$?

d-Separation

D-Separation kann verwendet werden, um Unabhängigkeiten in BNen zu finden:

Theorem (Soundness)

Für alle BN $N = (G, \Theta)$ und alle disjunkten Knotenmengen X, Y, Z gilt: $\mathsf{dsep}_G(X, Z, Y)$ impliziert $I_{\mathsf{Pr}_N}(X, Z, Y)$.

Den (nicht unsubtilen) Beweis lassen wir weg: wenn $\mathsf{desg}_G(X, Z, Y)$, dann $I_{\mathsf{Pr}_N}(X, Z, Y)$ aus den Graphoid Axiomen herleitbar.

Folgende naive Version von Vollständigkeit gilt offensichtlich nicht:

Für alle BN $N = (G, \Theta)$ und alle disjunkten Knotenmengen X, Y, Z gilt: $I_{\mathsf{Pr}_N}(X, Z, Y)$ impliziert $\mathsf{dsep}_G(X, Z, Y)$.

d-Separation

Es gilt aber Vollständigkeit in folgendem schwächeren Sinne:

Theorem (Vollständigkeit)

Für jeden DAG G gibt es KWTs Θ so dass für $N = (G, \Theta)$ gilt:
$I_{\mathsf{Pr}_N}(X, Z, Y)$ impliziert $\mathsf{dsep}_G(X, Z, Y)$ für alle disjunkten X, Y, Z.

Es folgt, dass man d-Separation nicht verbessern kann: kein nur auf G beruhender Test kann mehr Unabhängigkeiten herleiten.

Auch diesen Beweis lassen wir weg.

Naives Anwenden von d-Separierung erfordert das Prüfen von (schlimmstenfalls) 2^n Pfaden. Geht es auch besser?

d-Separation

Theorem

$\mathsf{dsep}_G(X, Z, Y)$ gdw. es keinen Pfad zwischen X und Y im Graph G' gibt, der aus G wie folgt entsteht:

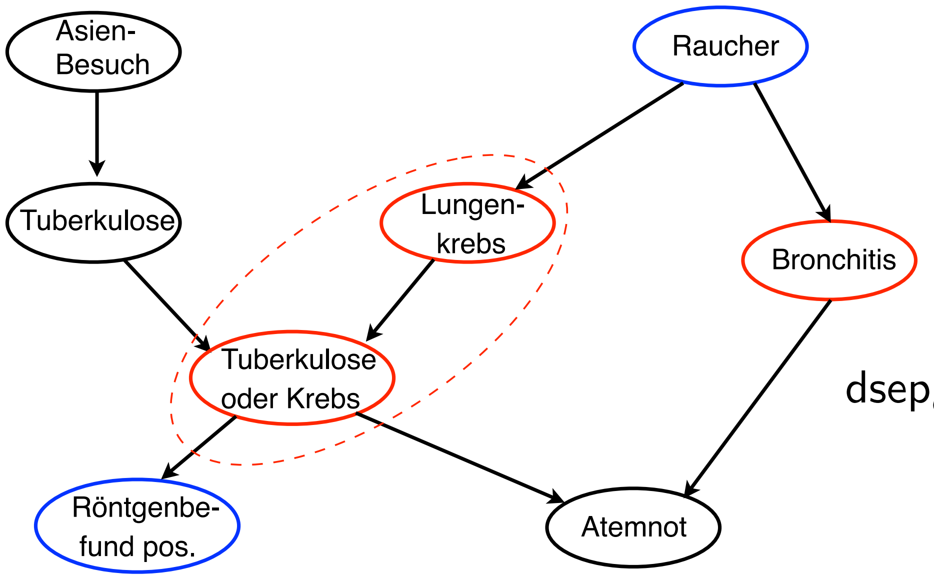
- entferne alle Blätter, die nicht zu $X \cup Y \cup Z$ gehören; wiederhole diesen Schritt erschöpfend;
- lösche alle Kanten, die an Knoten in Z beginnen.

d-Separation

Theorem

$\text{dsep}_G(X, Z, Y)$ gdw. es keinen Pfad zwischen X und Y im Graph G' gibt, der aus G wie folgt entsteht:

- entferne alle Blätter, die nicht zu $X \cup Y \cup Z$ gehören; wiederhole diesen Schritt erschöpfend;
- lösche alle Kanten, die an Knoten in Z beginnen.



$\text{dsep}_G($Tuberkulose, Lungenkrebs, Raucher, Röntgen, Bronchitis$)$?

d-Separation

Theorem

$\mathsf{dsep}_G(X, Z, Y)$ gdw. es keinen Pfad zwischen X und Y im Graph G' gibt, der aus G wie folgt entsteht:

- entferne alle Blätter, die nicht zu $X \cup Y \cup Z$ gehören; wiederhole diesen Schritt erschöpfend;
- lösche alle Kanten, die an Knoten in Z beginnen.

Komplexität:

- Erreichbarkeit in Graphen ist in Linearzeit entscheidbar
- Die entsprechenden Algorithmen können leicht angepasst werden, um auch dsep selbst in Linearzeit zu entscheiden.

Übersicht Vorlesung

- Motivation und Grundlagen
- Kapitel 1: Bayes-Netze
- Kapitel 2: Schlussfolgerungsmechanismen
- Kapitel 3: Komplexität und Approximation
- Kapitel 4: Maschinelles Lernen