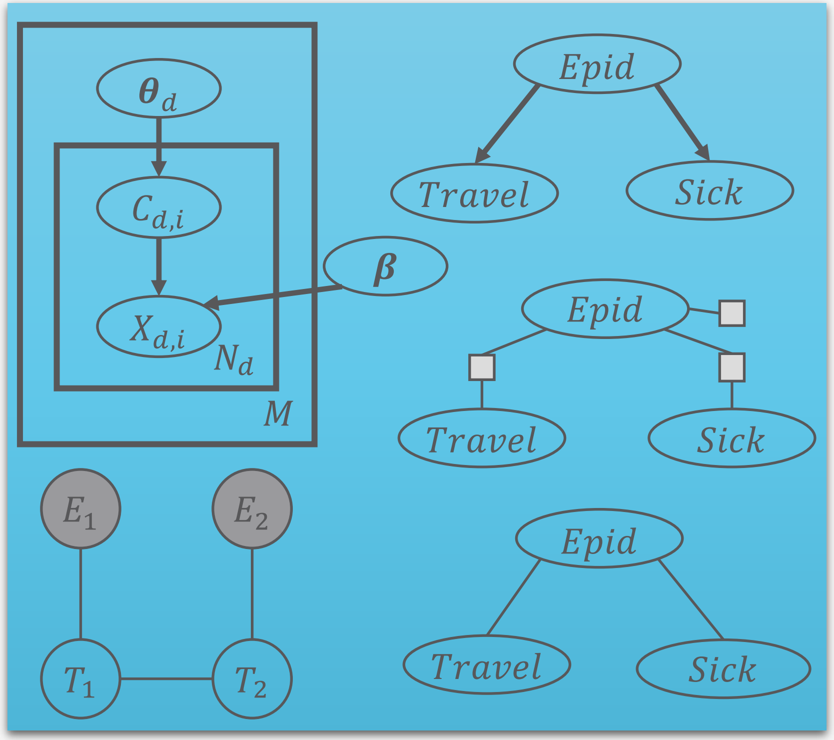
Episodische PGMs

Einführung in die Künstliche Intelligenz

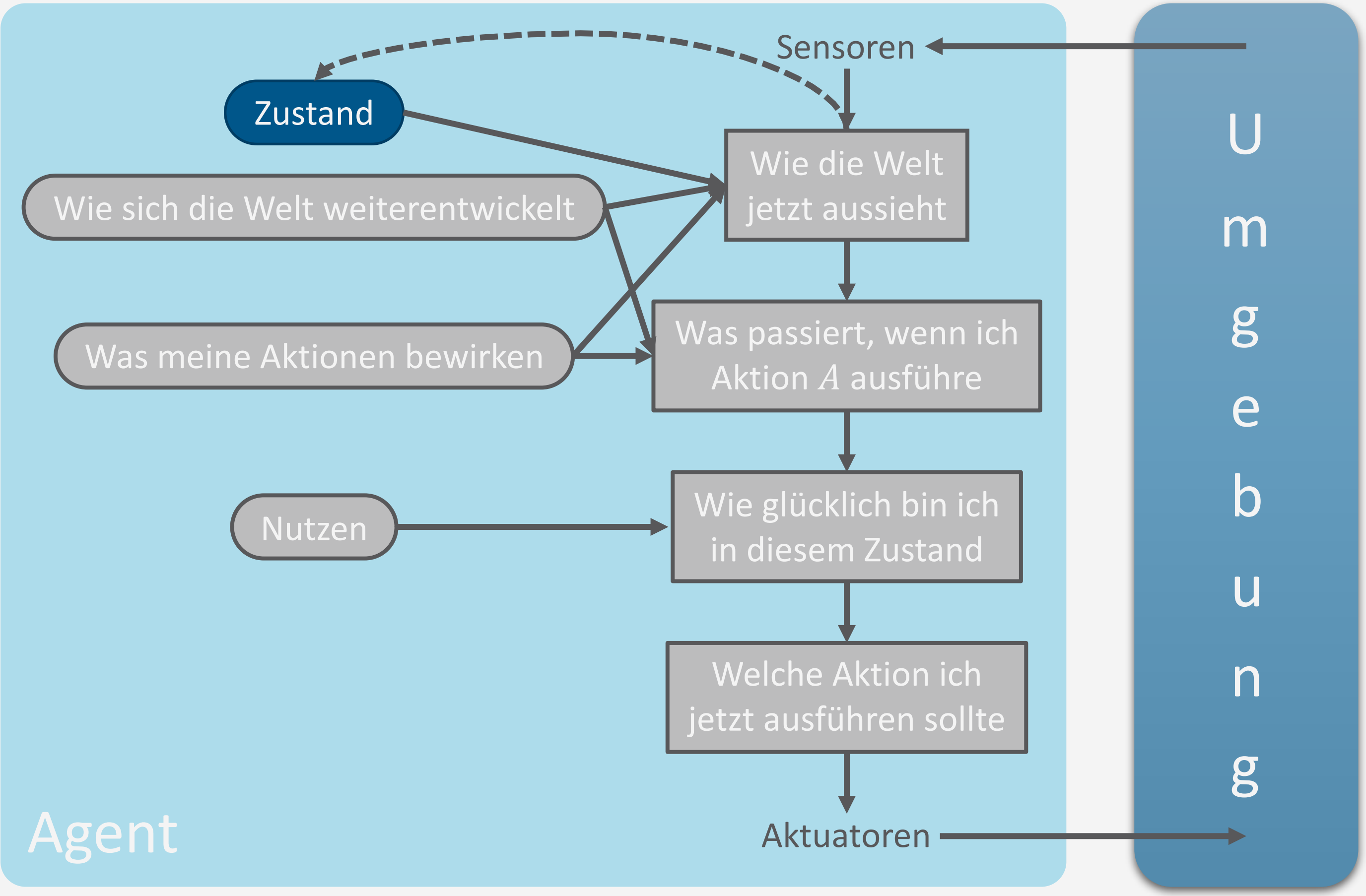


Tanya Braun
Arbeitsgruppe Data Science, Institut für Informatik

Inhalte

Einordnung der Vorlesung: *Modell- und nutzenbasierter Agent*

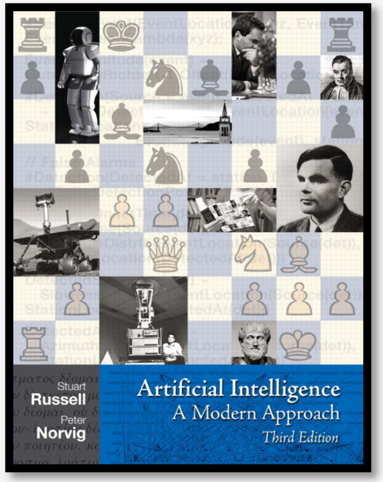
- Nachfolgende Themen der Vorlesung
  2. Episodische PGMs
  3. Exakte Inferenz in episodischen PGMs
  4. Approximative Inferenz in episodischen PGMs
  5. Lernalgorithmen für episodische PGMs
  6. Sequentielle PGMs und Inferenz
  7. Entscheidungstheoretische PGMs

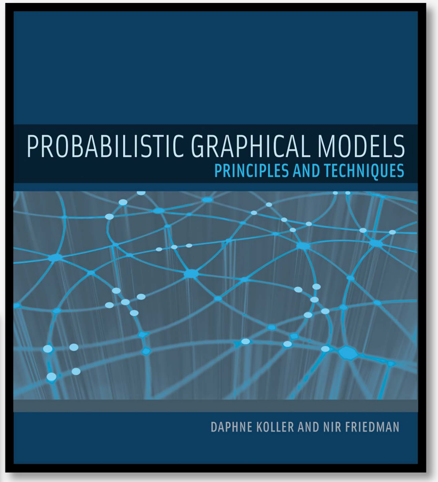


Literaturhinweise

Inhalte dieses Themenblocks werden in den folgenden Kapiteln der Vorlesungsbücher behandelt

- AIMA(de)
  - (Kap. 13: Grundlagen Wahrscheinlichkeitstheorie)
  - Kap. 14.1-2: Bayes Netze
- PGM: Ausschnitte der folgenden Abschnitte
  - (Kap. 2.1: Grundlagen Wahrscheinlichkeitstheorie)
  - Kap. 3.1-3.3: Bayes Netze
  - Kap. 4.2, 4.3, 4.4.1.1., 4.5: Ungerichtete Modelle





Die PGM-Abschnitte beinhalten Beispiele und Beweise, die nicht in die Folien eingeflossen sind.

- Wer also noch ein weiteres Beispiel sehen oder tiefer in die Beweise einsteigen möchte, sei auf diese Abschnitte verwiesen (inkl. 4.1: Beispiel)

Überblick: 2. Episodische PGMs

A. *Probabilistische Modellierung*
- Zufallsvariablen, vollständige gemeinsame Wahrscheinlichkeitsverteilung, Speicherkomplexität
- Inferenzaufgaben, Komplexität

B. *Gerichtete Modelle: Bayes Netze (BNs)*
- (Bedingte) Unabhängigkeiten, Faktorisierung
- Syntax, Semantik, graphische Darstellung, Speicherkomplexität

C. *Ungerichtete Modelle: Faktormodelle*
- Syntax, Semantik, graphische Darstellungen und deren Unterschiede, Speicherkomplexität
- Umwandlung von BNs zu Faktormodellen und deren graphischen Darstellungen

D. *Unabhängigkeiten in PGMs*
- Lokale, globale und paarweise Unabhängigkeiten
- Äquivalenzbedingungen von Faktormodellen und BNs

Beispielszenario

- Spielzeug-Beispiel: Grippe-Epidemie, gekennzeichnet durch
  - Epidemie herrscht (oder nicht) → $Epid$
  - Man reist durch die Gegend (oder nicht) → $Travel$
  - Man ist krank (oder nicht) → $Sick$
- Wahrscheinlichkeiten für die unterschiedlichen möglichen Kombinationen, z.B.
  - Es gibt keine Epidemie, man ist nicht krank und man reist nicht: 0.20
  - Es gibt keine Epidemie, man ist krank und man reist nicht: 0.24

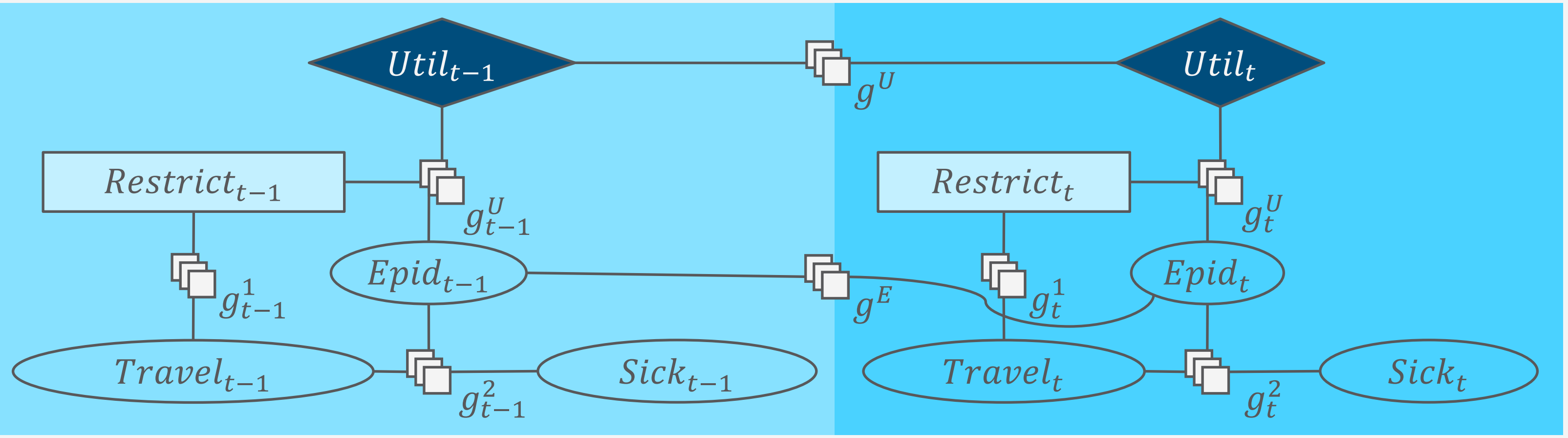
Beispiel dient der Illustration → Es hält keiner genaueren Prüfung stand!
- Ins Besondere werden wir im Laufe der Vorlesung einige Annahmen über dieses Szenario machen, welche in der realen Welt nicht zwangsweise unter Berücksichtigung aller Faktoren halten müssen.

$Epid$ $Travel$ $Sick$

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

Beispielszenario

- Wird im Laufe der Vorlesung weiter ausgebaut
  - Ziel: Umsetzung eines nutzenbasierten Agenten
    - Entwicklung über die Zeit: Epidemie-Verhalten von $t-1$ zu t
    - Entscheidungen zu Aktionen: Mobilität einschränken → $Restrict$
    - Nutzenfunktion: Ergebnis in $Util$
    - Ausblick auf Vorlesung zu entscheidungstheoretischen temporalen PGMs:



$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

Zufallsvariablen

- Beschreibung eines Szenarios mittels einer Menge von Zufallsvariablen
  - $\boldsymbol{R} = \{R_1, \dots, R_n\}$
  - In graphischen Modellen häufig als Ellipsen dargestellt
  - Beispiel

$$\{Epid, Travel, Sick\}$$

- Mögliche Werte, die eine Zufallsvariable annehmen kann = Domäne
  - $\mathrm{Val}(R) = \{v_1, \dots, v_m\}$
  - Wenn $|Val(R)| = 2$, dann Boolesche Domäne
  - Gegeben eine Menge $\boldsymbol{R}$ oder Sequenz $\mathcal{R}$: Kreuzprodukt der Domänen
  - Beispiel:

$$\mathrm{Val}(Epid) = \mathrm{Val}(Travel) = \mathrm{Val}(Sick) = \{true, false\}$$

$$Val(Epid, Travel) = \{(true, true), (true, false), (false, true), (false, false)\}$$

Notation:
- Variable: Anfang groß R
- Wert: Anfang klein r
- Menge: **fett gesetzt**
  $\boldsymbol{R} = \{R_1, \dots, R_n\}, \boldsymbol{r} = \{r_1, \dots, r_n\}$
- Sequenz: *kalligraphisch gesetzt*
  $\mathcal{R} = (R_1, \dots, R_n), \mathcal{r} = (r_1, \dots, r_n)$

Events

- Einen bestimmten Wert einer Zufallsvariable beobachten bzw. einer Zufallsvariable einen bestimmten Wert zuweisen = Event
  - $R = r$
    - $r \in \mathrm{Val}(R)$
  - Kurzschrift: r anstatt $R = r$, wenn R durch Kontext klar
    - Wenn $\mathrm{Val}(R)$ Boolesch, r für $R = true$ und $\neg r$ für $R = false$
  - Beispiel $\quad Epid = true \quad Epid = false$

    $\quad epid \quad \neg epid$
- Einer Menge von Zufallsvariablen $\boldsymbol{R}$ jeweils einen Wert der jeweiligen Domäne zuweisen = zusammengesetztes Event
  - $\{R = r\}_{R \in \boldsymbol{R}}$

Vollständige gemeinsame Wahrscheinlichkeitsverteilung

- Welt ω = zusammengesetztes Event für $\boldsymbol{R}$, i.e., $\boldsymbol{r}$
  - Beispiel:

$$epid, \neg travel, \neg\, sick$$

- Wahrscheinlichkeit für das Auftreten dieser Welt angeben
  - Beispiel

$$P(epid, \neg travel, \neg sick) = 0.05$$

Wie groß ist l?

- Vollständige gemeinsame Wahrscheinlichkeitsverteilung $P_{\boldsymbol{R}} = P(\boldsymbol{R})$ über alle (l) möglichen Welten
  - Eigentlich eine Funktion P_R : $\mathrm{Val}(\boldsymbol{R}) \rightarrow \mathbb{R}$ mit $\sum_{i=1}^{l} P(\omega_i) = 1$
    - ω_i: zusammengesetztes Event für $\boldsymbol{R}$
  - Auch bekannt als multivariate Verteilung, engl. *full joint (probability distribution)*

$Epid$ $Travel$ $Sick$

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

Speicherkomplexität

- Gegeben eine vollständige gemeinsame Wahrscheinlichkeitsverteilung $P_{\boldsymbol{R}}$ über alle (l) möglichen Welten einer Menge von Zufallsvariablen $\boldsymbol{R}$
- Speicherkomplexität: $O(r^n)$
  - $r = \max_{R \in \boldsymbol{R}} |\mathrm{Val}(R)|$
  - $n = |\boldsymbol{R}|$
  - Herleitung

$$\underbrace{\prod_{R \in \boldsymbol{R}} |\mathrm{Val}(R)|}_{\text{Genaue Größe}} \leq \prod_{R \in \boldsymbol{R}} \max_{R \in \boldsymbol{R}} |\mathrm{Val}(R)| = \prod_{R \in \boldsymbol{R}} r = r^{|\boldsymbol{R}|} = r^n$$

Exponentiell in n!

$Epid$ $Travel$ $Sick$

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

Nebenbemerkung: Bernoulli-Verteilung

- Diskrete Boolesche Zufallsvariable R
  - $\mathrm{Val}(R) = \{true, false\} = \{1,0\}$
  - Wahrscheinlichkeitsverteilung P_R
    - $P_R(R = 1) = p, P_R(R = 0) = 1 - p$
- Als Bernoulli-Verteilung mit Parameter $p = P_R(R = 1)$:

$$P(R = r \mid p) = \begin{cases} p^r(1-p)^{1-r} & falls\ r \in \{0,1\} \\ 0 & sonst. \end{cases}$$

  - Auch Null-Eins-Verteilung genannt
  - Auch notiert als $R \sim Ber(p)$

- Beispiel:
  - $Epid$ mit einer Verteilung:

$Epid$	P
$false$	0.8
$true$	0.2

  - $P(Epid = 1) = 0.2 = p$
  - Bernoulli-Verteilung:

$$P(Epid = r|0.2) = \begin{cases} 0.2^r 0.8^{1-r} & falls\ r \in \{0,1\} \\ 0 & sonst. \end{cases}$$

    - $P(R = 0|0.2) = 0.2^0 0.8^{1-0} = 1 \cdot 0.8^1 = 0.8$
    - $P(R = 1|0.2) = 0.2^1 0.8^{1-1} = 0.2^1 \cdot 1 = 0.2$

Nebenbemerkung: Multinomial-Verteilung bei $n = 1$

Auch Kategorialverteilung genannt

- Diskrete Zufallsvariable R mit m Ausprägungen, i.e., $\text{Val}(R) = \{r_1, \dots, r_m\}$
  - Wahrscheinlichkeitsverteilung P_R
  - $\boldsymbol{\pi}$ = m-dimensionaler Vektor mit Einträgen aus P_R
  - Wert r_i kodiert als *One-hot-Vektor* $\boldsymbol{v}$ über $\text{Val}(R)$ mit 1 an Position von r_i und 0 sonst
- Verteilung:

$$P(\boldsymbol{v}|\boldsymbol{\pi}) = \begin{cases} \prod_{i=1}^{m} \pi_i^{v_i} & falls\ r_i \in \text{Val}(R), wo\ v_i = 1 \\ 0 & sonst. \end{cases}$$

  - Auch notiert als $R \sim Mul(\boldsymbol{\pi})$

- Beispiel
  - $R = Topic$ mit $m = 4$:
    - Mögliche Themen (Topics) eines Dokuments
    - P_{Topic} Auftretenswahrscheinlichkeit
  - $\boldsymbol{\pi} = \langle 0.4, 0.3, 0.2, 0.1 \rangle$
  - $Topic = law \rightarrow \boldsymbol{v} = \langle 0,0,1,0 \rangle$
  - $P(\boldsymbol{v}|\boldsymbol{\pi}) = 0.4^0 \cdot 0.3^0 \cdot 0.2^1 \cdot 0.1^0 = 0.2$

$Topic$	P
$sport$	0.4
$econ$	0.3
law	0.2
cs	0.1

- Bei $m = 2$:
  - Bernoulli-Verteilung mit $\pi_1 = p$
  - $P(\boldsymbol{v}|\boldsymbol{\pi}) = \pi_1^{v_1}\pi_2^{v_2} = \pi_1^r(1 - \pi_1)^{1-r}$
    - Für $v = \langle 0,1 \rangle$ und $v = \langle 1,0 \rangle$, i.e., $r \in \{0,1\}$

Inferenzaufgaben

- Anfragenbeantwortungsproblem
  - Berechne die Antwort auf eine Anfrage gegeben einer vollständigen gemeinsamen Wahrscheinlichkeitsverteilung $P_{\boldsymbol{R}}$
    - Anfragen zu einer marginalen (bedingten) Wahrscheinlichkeit (-sverteilung)
    - Anfragetypen:
      - Marginale Wahrscheinlichkeit von Events
      - Marginale Wahrscheinlichkeitsverteilung von Zufallsvariablen
      - Marginale bedingte Wahrscheinlichkeit von Events gegeben Events (Evidenz / Beobachtungen)
      - Marginale bedingte Wahrscheinlichkeit von Zufallsvariablen gegeben Events (Evidenz / Beobachtungen)
- Nächste Folien
  - Syntax von Anfragen
  - Lösen einer Instanz es Anfragebeantwortungsproblems
    - Vorschau: Eliminiere alle Nicht-Anfrage-Variablen

Marginalanfragen

- Anfrage zu einer marginale Wahrscheinlichkeit (-sverteilung) an $P_{\boldsymbol{R}}$ über Zufallsvariablen $\boldsymbol{R}$
  - $P(\boldsymbol{S})$
    - $\mathrm{rv}(\boldsymbol{S}) \subseteq \boldsymbol{R}$
      - $\mathrm{rv}(.)$: Ausdruck, der sich auf die Zufallsvariablen der Eingabe bezieht
  - $\boldsymbol{S}$: Menge von
    - Zufallsvariablen (Anfrage zu Verteilung) oder
    - Events (Anfrage zur Wahrscheinlichkeit)
- Z.B.,

$$P(epid)$$
$$P(Epid, Travel)$$

- Wie beantworten wir eine Anfrage an $P_{\boldsymbol{R}}$?

$Epid$ $Travel$ $Sick$

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

Beantworten von Marginalanfragen

- Gegeben eine Anfrage $P(\boldsymbol{S})$ an $P_{\boldsymbol{R}}$ über Zufallsvariablen $\boldsymbol{R}$
- Eliminiere alle Nicht-Anfragevariablen $\boldsymbol{U} = \boldsymbol{R} \setminus \mathrm{rv}(\boldsymbol{S})$
  - Eliminieren = Aussummieren
    - Gegeben $R_1, \dots, R_m \in \boldsymbol{U}$:

$$P(\boldsymbol{S}) = \sum_{v_1 \in \mathrm{Val}(R_1)} \dots \sum_{v_m \in \mathrm{Val}(R_m)} P_{\boldsymbol{R}}(R_1 = v_1, \dots, R_m = v_m, \boldsymbol{S})$$

      - Für *jede* Wertekombination $\boldsymbol{s}$ von $\boldsymbol{S}$, summiere die Wahrscheinlichkeiten über *alle* Wertekombinationen $\boldsymbol{u}$ von $\boldsymbol{U}$ auf, auf die $\boldsymbol{s} \cup \boldsymbol{u}$ abbilden
- Beispiel: $P(Epid, Travel), \boldsymbol{U} = \{Sick\}$

$$P(Epid, Travel) = \sum_{v \in \mathrm{Val}(Sick)} P_{\boldsymbol{R}}(Epid, Travel, Sick = v)$$

$Epid$ $Travel$ $Sick$

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

Beantworten von Marginalanfragen

$$P(Epid, Travel) = \sum_{v \in \mathrm{Val}(Sick)} P(Epid, Travel, Sick = v)$$

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

+ +
+ +

$Epid$	$Travel$	P
$false$	$false$	0.44
$false$	$true$	0.36
$true$	$false$	0.11
$true$	$true$	0.09

Beantworten von Marginalanfragen

- Gegeben eine Anfrage $P(\boldsymbol{S})$ an $P_{\boldsymbol{R}}$ über Zufallsvariablen $\boldsymbol{R}$
- Eliminiere alle Nicht-Anfragevariablen $\boldsymbol{U} = \boldsymbol{R} \setminus \mathrm{rv}(\boldsymbol{S})$
  - Eliminieren = Aussummieren
    - Gegeben $R_1, \dots, R_m \in \boldsymbol{U}$:

$$P(\boldsymbol{S}) = \sum_{v_1 \in \mathrm{Val}(R_1)} \dots \sum_{v_m \in \mathrm{Val}(R_m)} P_{\boldsymbol{R}}(R_1 = v_1, \dots, R_m = v_m, \boldsymbol{S})$$

- Für *jede* Wertekombination $\boldsymbol{s}$ von $\boldsymbol{S}$, summiere die Wahrscheinlichkeiten über *alle* Wertekombinationen $\boldsymbol{u}$ von $\boldsymbol{U}$ auf, auf die $\boldsymbol{s} \cup \boldsymbol{u}$ abbilden
- Wenn $\boldsymbol{S}$ aus Events besteht, dann berücksichtige nur die Fälle, in denen die Werte in P_F mit den Events in $\boldsymbol{S}$ übereinstimmen
- Beispiel: $P(epid), \boldsymbol{U} = \{Travel, Sick\}$

$$P(epid) = \sum_{v_t \in \mathrm{Val}(Travel)} \sum_{v_s \in \mathrm{Val}(Sick)} P(epid, Travel = v_t, Sick = v_s)$$

$Epid$ $Travel$ $Sick$

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

Beantworten von Marginalanfragen

$$P(epid) = \sum_{v_t \in \mathrm{Val}(Travel)} \sum_{v_s \in \mathrm{Val}(Sick)} P(epid, Travel = v_t, Sick = v_s)$$

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

+

$Epid$	$Travel$	P
$true$	$false$	0.11
$true$	$true$	0.09

+

$Epid$	P
$true$	0.20

Bedingte Anfragen

- Anfrage für eine bedingte (auch: *konditionale*) marginale Wahrscheinlichkeit (-sverteilung) an $P_{\boldsymbol{R}}$
  - $P(\boldsymbol{S}|\boldsymbol{T})$
    - $\mathrm{rv}(\boldsymbol{S}) \subseteq \boldsymbol{R}, \mathrm{rv}(\boldsymbol{T}) \subseteq \boldsymbol{R}$
    - $\boldsymbol{S} \cap \boldsymbol{T} = \emptyset$
  - $\boldsymbol{S}$: Zufallsvariablen oder Events (wie vorher)
  - $\boldsymbol{T}$: Zufallsvariablen oder Events (Wahrnehmungen / Beobachtungen, genannt Evidenz)
  - Verallgemeinerung der marginalen Anfrage: Bedingte Anfrage mit $\boldsymbol{T} = \emptyset$
- Beispiel:

$$P(Sick|Epid)$$
$$P(Epid|sick)$$

$Epid$ $Travel$ $Sick$

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

Beantworten von bedingten Anfragen

- Gegeben eine Anfrage $P(\boldsymbol{S} \mid \boldsymbol{T})$ an $P_{\boldsymbol{R}}$
- Definition bedingter Wahrscheinlichkeit: $P(\boldsymbol{S} \mid \boldsymbol{T}) = \frac{P(\boldsymbol{S}, \boldsymbol{T})}{P(\boldsymbol{T})}$
  - $P(\boldsymbol{T})$ Normalisierungskonstante
    - Auch Partitionsfunktion oder Zustandssumme genannt
    - Abkürzende Notation: $\alpha, \frac{1}{Z}$ oder weggelassen mit „∝“ anstatt „=“
    - Wird zu einer leeren Anfrage bei $\boldsymbol{T} = \emptyset$: $P(.) \rightarrow$ Alles aussummieren, ergibt 1 bei einer Wahrscheinlichkeitsverteilung als Grundlage
- Reduziert sich auf Berechnung zweier Marginalanfragen: $P(\boldsymbol{S}, \boldsymbol{T})$, $P(\boldsymbol{T})$
  - Und: Für $P(\boldsymbol{T})$ müssen aus $P(\boldsymbol{S}, \boldsymbol{T})$ nur noch die Variablen in $\boldsymbol{S}$ eliminiert werden

$\rightarrow$ Berechnung von $P(\boldsymbol{S}, \boldsymbol{T})$ mit anschließender Normalisierung pro $\boldsymbol{t}$

$Epid$ $Travel$ $Sick$

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

Beantworten von bedingten Anfragen

- Gegeben eine Anfrage $P(\boldsymbol{S} \mid \boldsymbol{T})$ an $P_{\boldsymbol{R}}$
- Eliminiere alle Nicht-Anfragevariablen $\boldsymbol{U} = \boldsymbol{R} \setminus \mathrm{rv}(\boldsymbol{S}, \boldsymbol{T})$ und normalisiere pro $\boldsymbol{t} \in \mathrm{Val}(\boldsymbol{T})$
  - Gegeben $R_1 \dots R_m \in \boldsymbol{U}$:

$$P(\boldsymbol{S} \mid \boldsymbol{T}) = \frac{1}{P(\boldsymbol{T})} \sum_{v_1 \in \mathrm{Val}(R_1)} \dots \sum_{v_n \in \mathrm{Val}(R_m)} P_{\boldsymbol{R}}(R_1 = v_1, \dots, R_m = v_m, \boldsymbol{S}, \boldsymbol{T})$$

- Beispiel: $P(Sick \mid Epid), \boldsymbol{U} = \{Travel\}, P(\boldsymbol{T}) = P(Epid)$

$$P(Sick \mid Epid) = \frac{1}{P(Epid)} \sum_{v_t \in \mathrm{Val}(Travel)} P_{\boldsymbol{R}}(Epid, Travel = v_t, Sick)$$

$Epid$ $Travel$ $Sick$

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

Beantworten von bedingten Anfragen

- Gegeben eine Anfrage $P(\boldsymbol{S} \mid \boldsymbol{T})$ an $P_{\boldsymbol{R}}$
- Eliminiere alle Nicht-Anfragevariablen $\boldsymbol{U} = \boldsymbol{R} \setminus \mathrm{rv}(\boldsymbol{S}, \boldsymbol{T})$ und normalisiere pro $\boldsymbol{t} \in \mathrm{Val}(\boldsymbol{T})$
  - Gegeben $R_1 \dots R_m \in \boldsymbol{U}$:

$$P(\boldsymbol{S} \mid \boldsymbol{T}) = \frac{1}{P(\boldsymbol{T})} \sum_{v_1 \in \mathrm{Val}(R_1)} \dots \sum_{v_n \in \mathrm{Val}(R_m)} P_{\boldsymbol{R}}(R_1 = v_1, \dots, R_m = v_m, \boldsymbol{S}, \boldsymbol{T})$$

  - Was ist, wenn $\boldsymbol{T}$ aus Events $R = r$ besteht?
    - $P(R = r) = 1$ und damit $P(R \neq r) = 0$
    - Beispiel: $P(Epid \mid sick)$
    - Wird über einen Vorgang namens Absorption behandelt

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

Beantworten von bedingten Anfragen

- Gegeben eine Anfrage $P(\boldsymbol{S} \mid \boldsymbol{T})$ an $P_{\boldsymbol{R}}$
- Allgemeines Vorgehen:
  1. Evidenz absorbieren (optional)
  2. Nicht-Anfragevariablen eliminieren
  3. Normalisieren (momentan optional)

- Das erste „optional“ hängt davon ab, ob $\boldsymbol{T}$ Events beinhaltet
- Das zweite „optional“ hängt davon ab, ob $\boldsymbol{T} = \emptyset$
  - Später werden wir auch mit allgemeinen Verteilungen arbeiten, die immer im letzten Schritt normalisiert werden müssen, um eine Wahrscheinlichkeitsverteilung herzustellen

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

1. Absorption

Vorgehen:

a. Bei $R \neq r$ Wahrscheinlichkeiten auf 0 setzen
b. Zeilen mit Wahrscheinlichkeit 0 fallen lassen
c. Evidenzvariable fallen lassen
   - Kein Informationsverlust bei b. + c.

$$P(Epid \mid sick) \boxed{\propto} \sum_{v_t \in \mathrm{Val}(Travel)} P(Epid, Travel = v_t, sick)$$

Proportional zu

$Epid$	$Travel$	$Sick$	P	P
~~$false$~~	~~$false$~~	~~$false$~~	~~0.20~~	~~0~~
$false$	$false$	$true$	0.24	0.24
~~$false$~~	~~$true$~~	~~$false$~~	~~0.28~~	~~0~~
$false$	$true$	$true$	0.08	0.08
~~$true$~~	~~$false$~~	~~$false$~~	~~0.05~~	~~0~~
$true$	$false$	$true$	0.06	0.06
~~$true$~~	~~$true$~~	~~$false$~~	~~0.07~~	~~0~~
$true$	$true$	$true$	0.02	0.02

$Epid$	$Travel$	~~$Sick$~~	P
$false$	$false$	~~$true$~~	0.24
$false$	$true$	~~$true$~~	0.08
$true$	$false$	~~$true$~~	0.06
$true$	$true$	~~$true$~~	0.02

$Epid$	$Travel$	P
$false$	$false$	0.24
$false$	$true$	0.08
$true$	$false$	0.06
$true$	$true$	0.02

Absorption mit Dimensionsreduktion:
≙ Selektion mit $R = r$ und Projektion auf $\boldsymbol{R} \setminus R$
→ Reduziert die Dimension der Verteilung um 1

2. Elimination & 3. Normalisierung

$$P(Epid \mid sick) \propto \sum_{v_t \in \mathrm{Val}(Travel)} P(Epid, Travel = v_t, sick)$$

$Epid$	$Travel$	P
$false$	$false$	0.24
$false$	$true$	0.08
$true$	$false$	0.06
$true$	$true$	0.02

+

$Epid$	P
$false$	0.32
$true$	0.08

$Epid$	P
$false$	$\frac{0.32}{0.32 + 0.08} = 0.8$
$true$	$\frac{0.08}{0.32 + 0.08} = 0.2$

normalisieren

Entspricht Normalisierung wirklich durch $P(T) = P(sick)$ teilen?

$$P(Epid \mid sick) \propto \sum_{v_t \in \mathrm{Val}(Travel)} P(Epid, Travel = v_t, sick)$$

$$P(sick) \propto \sum_{v_e \in \mathrm{Val}(Epid)} \sum_{v_t \in \mathrm{Val}(Travel)} P(Epid = v_e, Travel = v_t, sick)$$

Berechnung $P(sick)$:

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

\+

$Epid$	$Sick$	P
$false$	$true$	0.32
$true$	$true$	0.08

\+

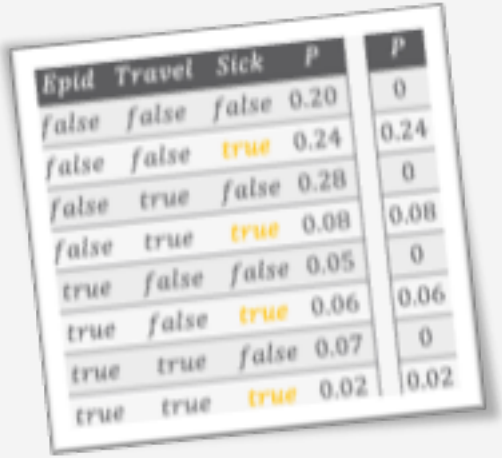
$Sick$	P
$true$	0.40

Normalisierungsschritt auf der vorherigen Folie

Absorption: Alternative Interpretation

Multiplikation von Faktoren/Verteilungen
≙ Join (⋈) über die Zufallsvariablen, Multiplikation der Wahrscheinlichkeiten

1. Bei $R \neq r$ Wahrscheinlichkeiten auf 0 setzen
   - ≙ Multiplikation mit einer $\langle 1,0 \rangle$-Verteilung über R, wobei $R = r$ auf 1 und $R \neq r$ auf 0 abbildet
     - Evidenzfaktor ϕ_e



Epid	Travel	Sick	P	P
false	false	false	0.20	0
false	false	true	0.24	0.24
false	true	false	0.28	0
false	true	true	0.08	0.08
true	false	false	0.05	0
true	false	true	0.06	0.06
true	true	false	0.07	0
true	true	true	0.02	0.02

Wahrscheinlichkeiten auf 0 setzen von vorheriger Folie

≙

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

$\cdot$

$Sick$	ϕ_e
$false$	0
$true$	1

⋮

$=$

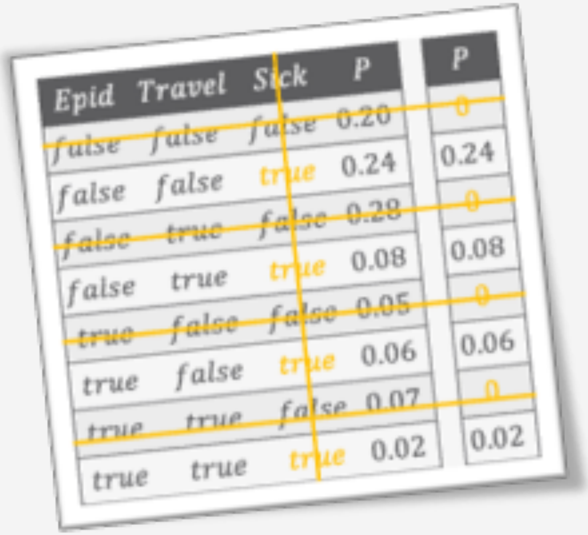
$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	$0.20 \cdot 0 = 0$
$false$	$false$	$true$	$0.24 \cdot 1 = 0.24$
$false$	$true$	$false$	$0.28 \cdot 0 = 0$
$false$	$true$	$true$	$0.08 \cdot 1 = 0.08$
$true$	$false$	$false$	$0.05 \cdot 0 = 0$
$true$	$false$	$true$	$0.06 \cdot 1 = 0.06$
$true$	$true$	$false$	$0.07 \cdot 0 = 0$
$true$	$true$	$true$	$0.02 \cdot 1 = 0.02$

⋮

Absorption: Alternative Interpretation

2. Zeilen mit Wahrscheinlichkeit 0 fallen lassen
3. Evidenzvariable fallen lassen

≙ Aussummieren der Evidenzvariable



Epid	Travel	Sick	P	P
false	false	false	0.20	0
false	false	true	0.24	0.24
false	true	false	0.28	0
false	true	true	0.08	0.08
true	false	false	0.05	0
true	false	true	0.06	0.06
true	true	false	0.07	0
true	true	true	0.02	0.02

Zeilen mit 0 und Spalte mit Evidenz fallenlassen

≙

Epid	Travel	Sick	P	P
$false$	$false$	$false$	0.20	0
$false$	$false$	$true$	0.24	0.24
$false$	$true$	$false$	0.28	0
$false$	$true$	$true$	0.08	0.08
$true$	$false$	$false$	0.05	0
$true$	$false$	$true$	0.06	0.06
$true$	$true$	$false$	0.07	0
$true$	$true$	$true$	0.02	0.02

Epid	Travel	P
+ $false$	$false$	0.24
+ $false$	$true$	0.08
+ $true$	$false$	0.06
+ $true$	$true$	0.02

Absorption: Alternative Interpretation

Absorption:
1. Bei $R \neq r$ Wahrscheinlichkeiten auf 0 setzen
2. Zeilen mit Wahrscheinlichkeit 0 fallen lassen
3. Evidenzvariable fallen lassen

≙ Multiplikation mit $\langle 1{,}0 \rangle$-Verteilung (ϕ_e)

≙ Aussummieren der Evidenzvariable

$$P(Epid \mid sick) \propto \sum_{v_t \in \mathrm{Val}(Travel)} P(Epid, Travel = v_t, sick)$$

$$P(Epid \mid sick) \propto \sum_{v_t \in \mathrm{Val}(Travel)} \sum_{v_s \in \mathrm{Val}(Sick)} P(Epid, Travel = v_t, Sick = v_s)\phi_e(Sick = v_s)$$

Laufzeitkomplexität

$$P(\boldsymbol{S} \mid \boldsymbol{T}) = \frac{1}{P(\boldsymbol{T})} \sum_{v_1 \in \mathrm{Val}(R_1)} \dots \sum_{v_n \in \mathrm{Val}(R_m)} P(R_1 = v_1, \dots, R_n = v_n, \boldsymbol{S}, \boldsymbol{T})$$

$Epid$ $Travel$ $Sick$

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

- Worst-case: Leere Anfrage $P(.)$
  - $|\boldsymbol{R}| = m = n$ Zufallsvariablen zu eliminieren durch Summe über $Val(\boldsymbol{R})$
- Laufzeitkomplexität: $O(r^n)$
  - $r = \max_{R \in \boldsymbol{R}} |\mathrm{Val}(R)|$
  - $n = |\boldsymbol{R}|$
  - Herleitung gleich zur Platzkomplexität

= Speicherkomplexität

Schwere des Problems

- Anfragebeantwortungsproblem als Entscheidungsproblem:
  - Gegeben eine vollständige gemeinsame Verteilung $P_{\boldsymbol{R}}$ über Zufallsvariablen $\boldsymbol{R}$, eine Zufallsvariable $R \in \boldsymbol{R}$, und eine Belegung $r \in \mathrm{Val}(R)$, entscheide ob $P(R = r) > 0$ ist
  - Entscheidungsproblem ist $\mathcal{NP}$-vollständig
    - [Beweis durch Reduktion auf 3-SAT-Problem]
- Anfragebeantwortungsproblem
  - Gegeben eine vollständige gemeinsame Verteilung $P_{\boldsymbol{R}}$ über Zufallsvariablen $\boldsymbol{R}$, eine Zufallsvariable $R \in \boldsymbol{R}$, und eine Belegung $r \in \mathrm{Val}(R)$, berechne $P(R = r)$
  - Anfragebeantwortungsproblem ist $\#\mathcal{P}$-vollständig
    - $\#\mathcal{P}$: Klasse von Problemen über „Wie viele Lösungen erfüllen bestimmte Anforderungen?“
    - [Totale Wahrscheinlichkeit von Graph-Instanziierungen, die konsistent mit $R = r$ sind, i.e., gewichtetes *Abzählen* von Instanziierungen, mit der Wahrscheinlichkeit der Instanziierung als Gewicht]

Zwischenzusammenfassung

- Zufallsvariablen, Events, Wahrscheinlichkeiten
- Vollständige gemeinsame Verteilung
  - Über eine Menge von Zufallsvariablen
  - Speicherkomplexität
- Inferenzaufgabe
  - Anfragebeantwortungsproblem
    - Anfragen für marginale (bedingte) Wahrscheinlichkeit (-sverteilung)
  - Vorgehen: Evidenz absorbieren, Nicht-Anfragevariablen eliminieren, normalisieren
    - Operationen:
      - Absorption als Selektion und Projektion
      - Produkt von Faktoren / Verteilungen als Join mit Multiplikation
      - Eliminieren / Aussummieren als Summe über die Werte der zu eliminierenden Zufallsvariable
  - Laufzeitkomplexität exponentiell in Anzahl der Variablen,
    - Schwere des Problems: $\mathcal{NP}$ bzw. $\#\mathcal{P}$-vollständig

Überblick: 2. Episodische PGMs

A. *Probabilistische Modellierung*
- Zufallsvariablen, vollständige gemeinsame Wahrscheinlichkeitsverteilung, Speicherkomplexität
- Inferenzaufgaben, Komplexität

B. ***Gerichtete Modelle: Bayes Netze (BNs)***
- (Bedingte) Unabhängigkeiten, Faktorisierung
- Syntax, Semantik, graphische Darstellung, Speicherkomplexität

C. *Ungerichtete Modelle: Faktormodelle*
- Syntax, Semantik, graphische Darstellungen und deren Unterschiede, Speicherkomplexität
- Umwandlung von BNs zu Faktormodellen und deren graphischen Darstellungen

D. *Unabhängigkeiten in PGMs*
- Lokale, globale und paarweise Unabhängigkeiten
- Äquivalenzbedingungen von Faktormodellen und BNs

Problem: r^n → Kombinatorische Explosion!

$Man.war$	$Man.virus$	$Nat.flood$	$Nat.fire$	$Treat.eve.m_1$	$Treat.eve.m_2$	$Epid$	$Travel.eve$	$Sick.eve$	P
false	false	false	false	false	false	false	false	false	0.025
false	false	false	false	false	false	false	false	true	0.009
false	false	false	false	false	false	false	true	false	0.012
false	false	false	false	false	false	false	true	true	0.005
false	false	false	false	false	false	true	false	false	0.017
false	false	false	false	false	false	true	false	true	0.028
false	false	false	false	false	false	true	true	false	0.003
false	false	false	false	false	false	true	true	true	0.001
false	false	false	false	false	true	false	false	false	0.025
false	false	false	false	false	true	false	false	true	0.009
false	false	false	false	false	true	false	true	false	0.012
false	false	false	false	false	true	false	true	true	0.005
false	false	false	false	false	true	true	false	false	0.017
false	false	false	false	false	true	true	false	true	0.028

9 Zufallsvariablen → $2^9 = 512$ mögliche Welten

(Bedingte) Unabhängigkeiten als Lösung?

- Unabhängigkeit
  - Idee: Zwei Events beeinflussen sich nicht gegenseitig
    - Wissen über das eine Event beeinflusst nicht die Wahrscheinlichkeit, dass das andere Event eintritt
  - Beispiel: Ausgänge zweier unterschiedlicher Münzwürfe
- Bedingte Unabhängigkeit
  - Idee: Zwei Events sind nicht unabhängig voneinander...
    - Wissen über das eine Event verändert die Wahrscheinlichkeit, dass das andere Event eintritt
  - ... aber haben einen gemeinsamen Einflussfaktor zwischen ihnen
    - Wenn der bekannt ist, dann hat das Wissen über das eine Event keinen Einfluss mehr auf die Wahrscheinlichkeit des anderen Event, da alles Wissenswerte bereits durch den Faktor dazwischen bekannt ist
  - Beispiel: Sonnenbrillenverkaufszahlen und Hautkrebsvorkommen
    - Einflussfaktor dazwischen: warmes Klima mit viel Sonnenschein

ACHTUNG: Bedingte Unabhängigkeiten erwecken manchmal den Eindruck einer Ursache-Wirkung-Beziehung, welche unter Umständen auch tatsächlich vorhanden ist. Das muss aber nicht so sein!
→ Stichwort Korrelation vs. Kausalität
→ Vorwärtszeiger:
Lernalgorithmen für episodische PGMs

(Bedingte) Unabhängigkeiten als Lösung?

- Unabhängigkeit
  - Formale Definition: Mengen von Zufallsvariablen $\boldsymbol{R}_1, \boldsymbol{R}_2$ sind unabhängig voneinander, wenn gilt:
    $$P(\boldsymbol{R}_1, \boldsymbol{R}_2) = P(\boldsymbol{R}_1) \cdot P(\boldsymbol{R}_2)$$
    - Notation: $\boldsymbol{R}_1 \perp \boldsymbol{R}_2$

- Bedingte Unabhängigkeit
  - Formale Definition: Mengen von Zufallsvariablen $\boldsymbol{R}_1, \boldsymbol{R}_2$ sind bedingt unabhängig voneinander gegeben eine Menge von Zufallsvariablen $\boldsymbol{R}'$, wenn gilt:
    $$P(\boldsymbol{R}_1, \boldsymbol{R}_2 \mid \boldsymbol{R}') = P(\boldsymbol{R}_1 \mid \boldsymbol{R}') \cdot P(\boldsymbol{R}_2 \mid \boldsymbol{R}')$$
    - Notation: $\boldsymbol{R}_1 \perp \boldsymbol{R}_2 \mid \boldsymbol{R}'$
    - Unabhängigkeit fällt unter bedingte Unabhängigkeit mit leerem $\boldsymbol{R}'$: $\boldsymbol{R}' = \emptyset \Rightarrow \boldsymbol{R}_1 \perp \boldsymbol{R}_2 = \boldsymbol{R}_1 \perp \boldsymbol{R}_2 \mid \boldsymbol{R}'$

(Bedingte) Unabhängigkeiten als Lösung?

- Unabhängigkeit: $P(\boldsymbol{R}_1, \boldsymbol{R}_2) = P(\boldsymbol{R}_1) \cdot P(\boldsymbol{R}_2)$
  - Beispiel: Gilt $Travel \perp Sick$?
    - $P(Travel, Sick) \stackrel{?}{=} P(Travel) \cdot P(Sick)$ → Hier nur der $true, true$ Fall:
      - $P(travel, sick) = 0.1$
      - $P(travel) = 0.45$
      - $P(sick) = 0.4$
      - $P(travel) \cdot P(sick) = 0.45 \cdot 0.4 = 0.18 \neq 0.1$ ✗
- Bedingte Unabhängigkeit: $P(\boldsymbol{R}_1, \boldsymbol{R}_2 | \boldsymbol{R}') = P(\boldsymbol{R}_1 | \boldsymbol{R}') \cdot P(\boldsymbol{R}_2 | \boldsymbol{R}')$
  - Beispiel: Gilt $Travel \perp Sick \mid Epid$?
    - $P(Travel, Sick | Epid) \stackrel{?}{=} P(Travel | Epid) \cdot P(Sick | Epid)$ ✗

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.20
$false$	$false$	$true$	0.24
$false$	$true$	$false$	0.28
$false$	$true$	$true$	0.08
$true$	$false$	$false$	0.05
$true$	$false$	$true$	0.06
$true$	$true$	$false$	0.07
$true$	$true$	$true$	0.02

(Bedingte) Unabhängigkeiten als Lösung?

- Gilt $Travel \perp Sick \mid Epid$ nun?
  - $P(Travel, Sick|Epid) \stackrel{?}{=} P(Travel|Epid) \cdot P(Sick|Epid)$ ✓

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.6375
$false$	$false$	$true$	0.1125
$false$	$true$	$false$	0.2125
$false$	$true$	$true$	0.0375
$true$	$false$	$false$	0.45
$true$	$false$	$true$	0.15
$true$	$true$	$false$	0.3
$true$	$true$	$true$	0.1

$Epid$	$Travel$	P
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

$Epid$	$Sick$	P
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

P
$0.75 \cdot 0.85 = 0.6375$
$0.75 \cdot 0.15 = 0.1125$
$0.25 \cdot 0.85 = 0.2125$
$0.25 \cdot 0.15 = 0.0375$
$0.6 \cdot 0.75 = 0.45$
$0.6 \cdot 0.25 = 0.15$
$0.4 \cdot 0.75 = 0.3$
$0.4 \cdot 0.25 = 0.1$

$Epid$ $Travel$ $Sick$

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.51
$false$	$false$	$true$	0.09
$false$	$true$	$false$	0.17
$false$	$true$	$true$	0.03
$true$	$false$	$false$	0.09
$true$	$false$	$true$	0.03
$true$	$true$	$false$	0.06
$true$	$true$	$true$	0.02

- Wie hilft uns das?

(Bedingte) Unabhängigkeiten als Lösung?

- Idee: Zerlege die vollständige gemeinsame Verteilung in ihre Bestandteile gemäß Unabhängigkeiten
  - Beispiel: $P(Epid, Travel, Sick) = P(Epid) \cdot P(Travel|Epid) \cdot P(Sick|Epid)$

$Epid$ $Travel$ $Sick$

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.51
$false$	$false$	$true$	0.09
$false$	$true$	$false$	0.17
$false$	$true$	$true$	0.03
$true$	$false$	$false$	0.09
$true$	$false$	$true$	0.03
$true$	$true$	$false$	0.06
$true$	$true$	$true$	0.02

$Epid$	P
$false$	0.8
$true$	0.2

$Epid$	$Travel$	P
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

$Epid$	$Sick$	P
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

$P(epid)$
0.2

$Epid$	$P(travel\|Epid)$
$false$	0.25
$true$	0.4

$Epid$	$P(sick\|Epid)$
$false$	0.15
$true$	0.25

Zugegeben: Bringt bei drei Zufallsvariablen nicht viel
- $2 + 4 + 4$ Einträge anstatt 8 Einträge
- *Da Wahrscheinlichkeiten auf* 1 *addieren:* $1 + 2 + 2$ *Einträge anstatt* 7 *Einträge genug*

..., aber bei steigender Zahl an Variablen schon

(Bedingte) Unabhängigkeiten als Lösung?

- Idee: Zerlege die vollständige gemeinsame Verteilung in ihre Bestandteile gemäß Unabhängigkeiten
  - Für die Anfragenbeantwortung nutzbar
  - *Intuition*:
    - Aussummieren von Nicht-Anfragevariablen aus einem Produkt
      - Es gilt das *Distributivgesetz* → Faktoren aus der inneren Summe soweit wie möglich nach vorn ziehen (ausklammern)
      - Aussummieren aus (kleineren) Teilprodukten
      - Grundidee der Variablenelimination (erster Algorithmus in Thema 3)
    - Beispiel: $\sum_{s \in \mathrm{Val}(Sick)} P(Epid, Travel, Sick = s) = \sum_{s \in \mathrm{Val}(Sick)} P(Epid) \cdot P(Travel|Epid) \cdot P(Sick = s|Epid)$

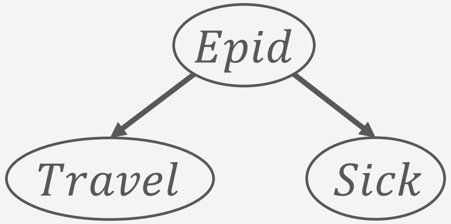
$$= P(Epid) \cdot P(Travel|Epid) \cdot \sum_{s \in \mathrm{Val}(Sick)} P(Sick = s|Epid)$$

Graphische Repräsentation der Zerlegung

- Explizite Repräsentation der Zerlegung von P_F durch probabilistisches graphisches Modell (PGM)
  - Zufallsvariablen als Knoten
  - Kanten kodieren bedingte Unabhängigkeiten $\boldsymbol{R}_1 \perp \boldsymbol{R}_2 | \boldsymbol{R}'$
    - $\boldsymbol{R}'$ als Elternknoten von $\boldsymbol{R}_1, \boldsymbol{R}_2$
  - Pro Knoten: $P\big(R|\mathrm{Pa}(R)\big)$
    - *AKA conditional probability distribution/table (CPD/CPT)*
    - $\mathrm{Pa}(R) = \emptyset$: a-priori Verteilung
    - $P_F = \prod_{R \in \boldsymbol{R}} P\big(R|\mathrm{Pa}(R)\big)$

  →Bayes Netz (BN)

$Epid$	P
$false$	0.8
$true$	0.2



$Epid$	$Travel$	P
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

$Epid$	$Sick$	P
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.51
$false$	$false$	$true$	0.09
$false$	$true$	$false$	0.17
$false$	$true$	$true$	0.03
$true$	$false$	$false$	0.09
$true$	$false$	$true$	0.03
$true$	$true$	$false$	0.06
$true$	$true$	$true$	0.02

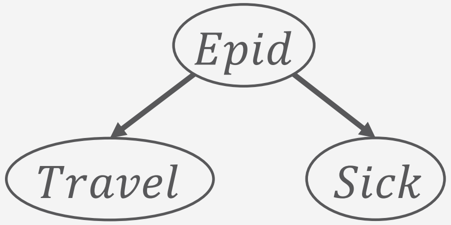
Bayes Netze (BN)

- Bayes Netz B: Gerichteter, azyklischer Graph $B = (V, E)$
  - Jedes $v \in V$ ist gelabelt mit einer CPD $P\big(v|Pa(v)\big)$
- Semantik
  - Jedes $v \in V$ steht für eine Zufallsvariable $R \in \boldsymbol{R}$
  - *Lesart 1*: B kodiert die <u>lokalen</u> Unabhängigkeiten
    - Für jede Zufallsvariable $R \in \boldsymbol{R}$: $\big(R \perp \mathrm{Ndesc}(R) | \mathrm{Pa}(R)\big)$
      - $\mathrm{Ndesc}(R)$: alle Nicht-Nachfahren von R, i.e., $\boldsymbol{R} \setminus \mathrm{Desc}(R)$
  - *Lesart 2*: B repräsentiert die vollständige gemeinsame Verteilung

$$P_B = P(\boldsymbol{R}) = \prod_{R \in \boldsymbol{R}} P\big(R|\mathrm{Pa}(R)\big)$$

    - Multiplikation von (bedingten) Wahrscheinlichkeitsverteilungen ergibt wieder eine Wahrscheinlichkeitsverteilung

$Epid$	P
$false$	0.8
$true$	0.2



$Epid$	$Travel$	P
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

$Epid$	$Sick$	P
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

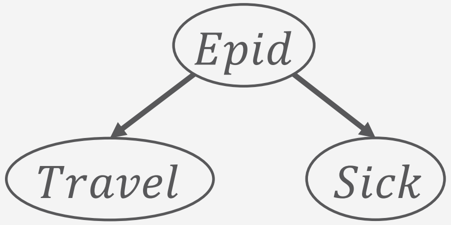
Speicherkomplexität

- Gegeben ein Bayes Netz B über Zufallsvariablen $\boldsymbol{R}$
- Speicherkomplexität: $O(n \cdot r^m)$
  - $n = |\boldsymbol{R}|$
  - $m = \max_{R \in \boldsymbol{R}} |\mathrm{Pa}(R)| + 1$ … Warum $+1$?
  - $r = \max_{R \in \boldsymbol{R}} |\mathrm{Val}(R)|$
  - Herleitung

$$\sum_{R \in \boldsymbol{R}} |\mathrm{Val}(R)|^{|\mathrm{Pa}(R)|+1} \leq \sum_{R \in \boldsymbol{R}} \overbrace{\max_{R \in \boldsymbol{R}} |\mathrm{Val}(R)|^{\max_{R \in \boldsymbol{R}} |\mathrm{Pa}(R)|+1}}^{\text{Größtmögliche CPD}} = \sum_{R \in \boldsymbol{R}} r^m = n \cdot r^m$$

  - Nicht mehr exponentiell in n, sondern in Anzahl Elternknoten m
    - Annahme: $m \ll n$, so dass $O(n \cdot r^m) \ll O(r^n)$

$Epid$	P
$false$	0.8
$true$	0.2



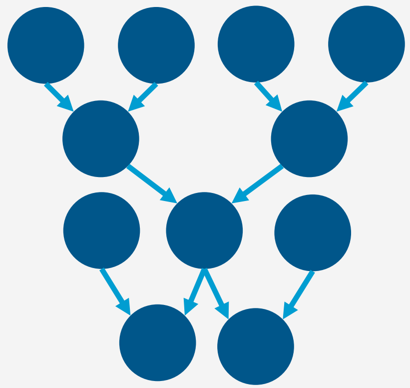
$Epid$	$Travel$	P
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

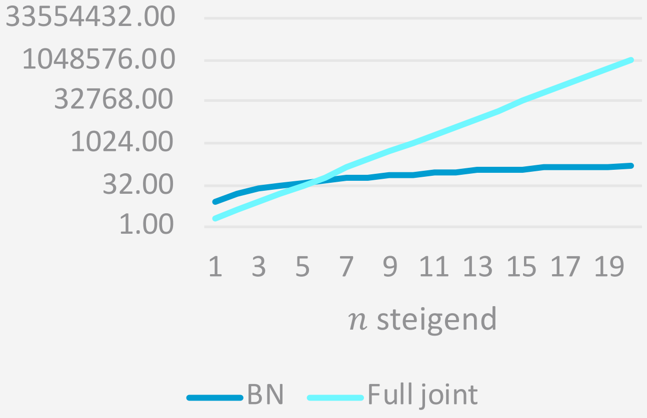
$Epid$	$Sick$	P
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

Beispiele Graphstrukturen

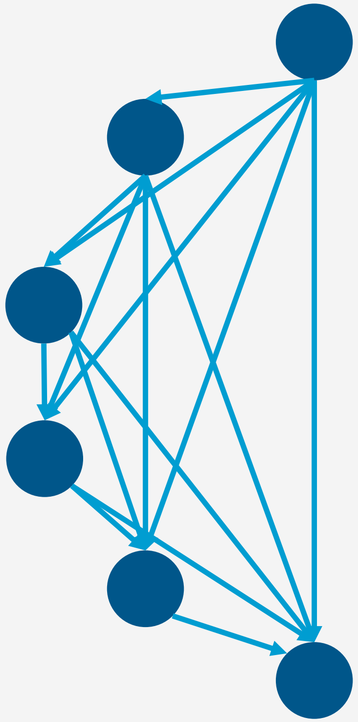
- Polytree BN mit $m = \max_{R \in \boldsymbol{R}} |\mathrm{Pa}(R)| + 1 = 3$
- BN mit $m = \max_{R \in \boldsymbol{R}} |\mathrm{Pa}(R)| + 1 = n$

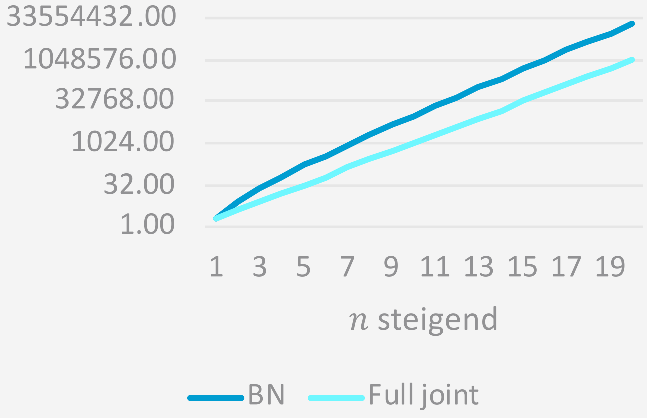
Was ist nochmal ein Polytree?





$O(n \cdot 2^2)$ vs. $O(2^n)$ mit $r = 2$



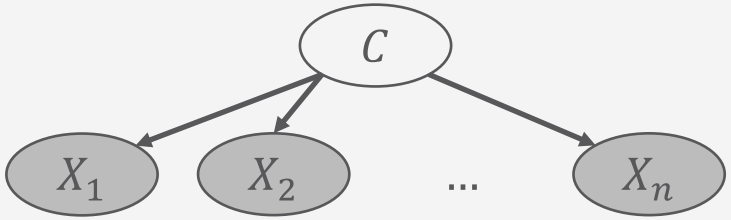


$O(n \cdot 2^n)$ vs. $O(2^n)$ mit $r = 2$

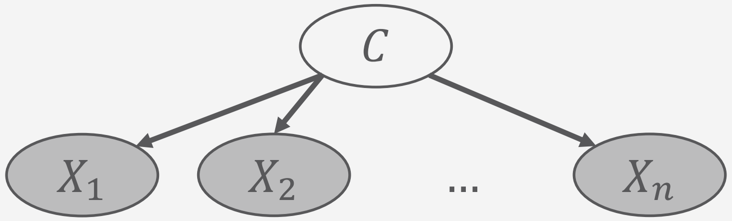
Anwendung von BNs: Naive Bayes Klassifizierer

Wenn in einem Modell bestimmte Zufallsvariablen immer mit Beobachtungen belegt werden, wie hier, dann werden die Knoten manchmal farblich gefüllt.

- Klassifikation:
  - Menge von Features $X_1, \dots, X_n$
  - Zuweisung eines Klassennamens c (Label) auf Basis der Featurewerte $x_1, \dots, x_n$
- Modellierung als BN:
  - Intuition: Klasse c beeinflusst die Ausprägung der Featurewerte $x_1, \dots, x_n$
  - $\boldsymbol{R} = \{X_1, \dots, X_n, C\}$ Menge von Zufallsvariablen
    - $X_1, \dots, X_n$ beobachtbar
    - C latent (nicht beobachtbar)
  - Faktorisierung: $P(C, X_1, \dots, X_n) = P(C) \prod_{i=1}^{n} P(X_i | C)$
  - Anfrage an BN: $P(C | x_1, \dots, x_n) \rightarrow$ Zuweisung des Klassennamens mit höchster Wahrscheinlichkeit, i.e., $\arg\max_{c \in Val(C)} P(c | x_1, \dots, x_n)$



Anwendung von BNs: Naive Bayes Klassifizierer

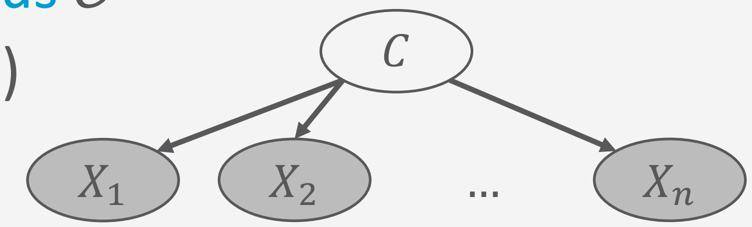


- Vorteile:
  - Einfach
  - Relativ wenige Parameter (Einträge in den CPDs) zu lernen
  - Braucht daher wenig Daten
    - Vergleich dazu tiefe neuronale Netze: Brauchen sehr viele Daten
- Früher durchaus für medizinische Diagnostik genutzt
  - Genauigkeit leidet unter starken Annahmen zur bedingten Unabhängigkeit zwischen den Features
- Idee aber weiterhin genutzt, z.B. in der Textklassifizierung – Topic Modellierung
  - Dokument einer von mehreren Kategorien (Topics/Themen = Klassenlabels) zuordnen
  - *Document Retrieval (DR)*: Gegeben eines Eingabedokuments, Dokumente gleichen Topics ausgeben

David M. Blei, Andrew Y. Ng, and Michael I. Jordan: *Latent Dirichlet Allocation*. Journal of Machine Learning Research, 2003.
Mehr zu Topic Modellierung auch in Box 17.E (Kap. 17) im PGM Buch. Achtung: Die Modelle dort sind subtil anders definiert.

Topic Modellierung

- Gegeben: Dokumente, zusammengesetzt aus Worten = Korpus $\mathcal{C}$
- Gesucht: Topic(-zusammensetzung) dieser Dokument (latent)
- Annahmen
  - Bag-of-Words: Dokument = Menge von Worten
    - Reihenfolge der Worte ignorieren
    - Standard Vorverarbeitungsschritte
      - Eliminierung von so genannten *stop words* (“and“, “the“, “a“, ...)
      - So genanntes *stemming*: Zurückführung aller Worte auf ein Grundwort / Wortstamm (“used“ → “use“, “running“ → “run“, ...)
      - Es ergibt sich quasi ein Standard-Lexikon von Stammworten, aus denen sich Dokumente zusammensetzen
  - Vorkommen von Worten hängt vom Topic ab
    - Beispiel: „Torwart“ wird in Dokumenten mit dem Thema *Sport* eher vorkommen als in Dokumenten mit dem Thema *Wirtschaft*

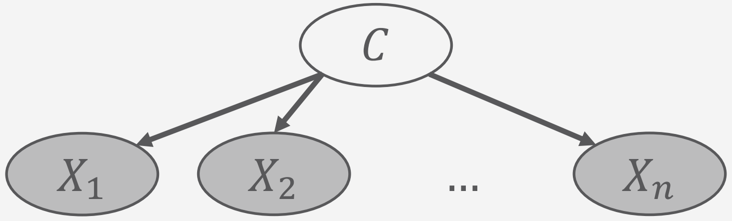


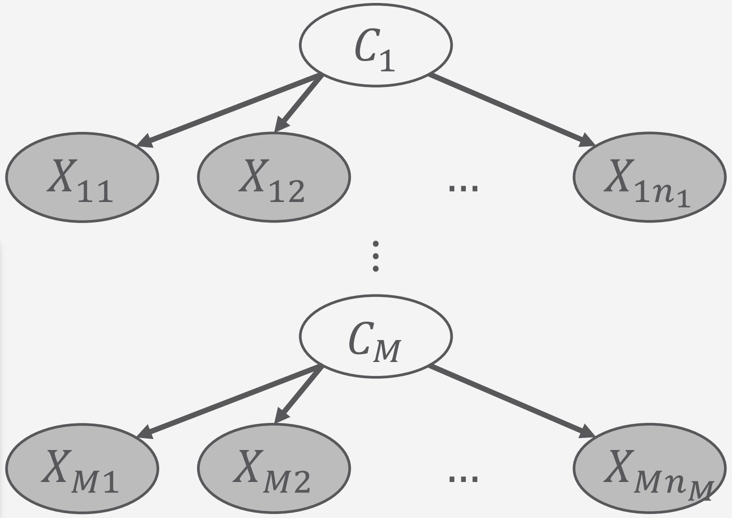
David M. Blei, Andrew Y. Ng, and Michael I. Jordan: *Latent Dirichlet Allocation*. Journal of Machine Learning Research, 2003.

Topic Modellierung mittels Naive Bayes: Mixture of Unigrams

- Weitere Annahme: Jedes Dokument hat genau <u>ein</u> Topic k
  - Topic eigentlich nur ein Index (kein Name)
- Pro Dokument $d \in \mathcal{C}$: Zufallsvariable X_i bezeichnet, welches Wort w aus einem Lexikon $\mathcal{D}$ an Position i in d vorkommt
  - I.e., $\mathrm{Val}(X_i) = \{w\}_{w \in \mathcal{D}}$
    - Bzw. $\mathrm{Val}(X_i) = \mathcal{D}$ mit $\mathcal{D}$ als Menge der Wörter im Lexikon
  - X_i unabhängig von X_j gegeben k:
    $$P(X_i, X_j | C) = P(X_i | C) P(X_j | C)$$
  - Annahme: Auftreten von w unabhängig von i:
    $$P(X_i = w) = P(X_j = w)$$
  - Jedes C_d hat die gleiche A-priori-Verteilung $P(C)$:
    $$P(C_d) = P(C_{d'}) = P(C)$$
  - Jedes X_{di} hat die gleiche CPD $P(X_i | C)$:
    $$P(X_{di} | C_d) = P(X_{d'i} | C_{d'}) = P(X_i | C)$$

parameter sharing: Mehrfachnutzung einer Verteilung (gleiche Multinomial-*parameter* $\boldsymbol{\pi}$)

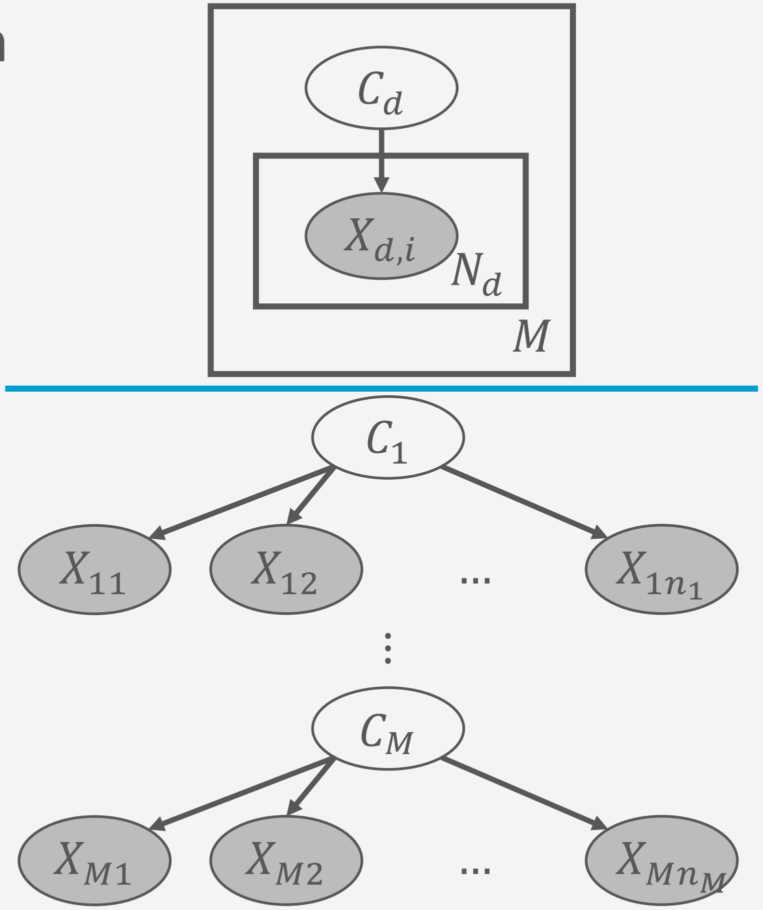




David M. Blei, Andrew Y. Ng, and Michael I. Jordan: *Latent Dirichlet Allocation*. Journal of Machine Learning Research, 2003.

Plate Notation

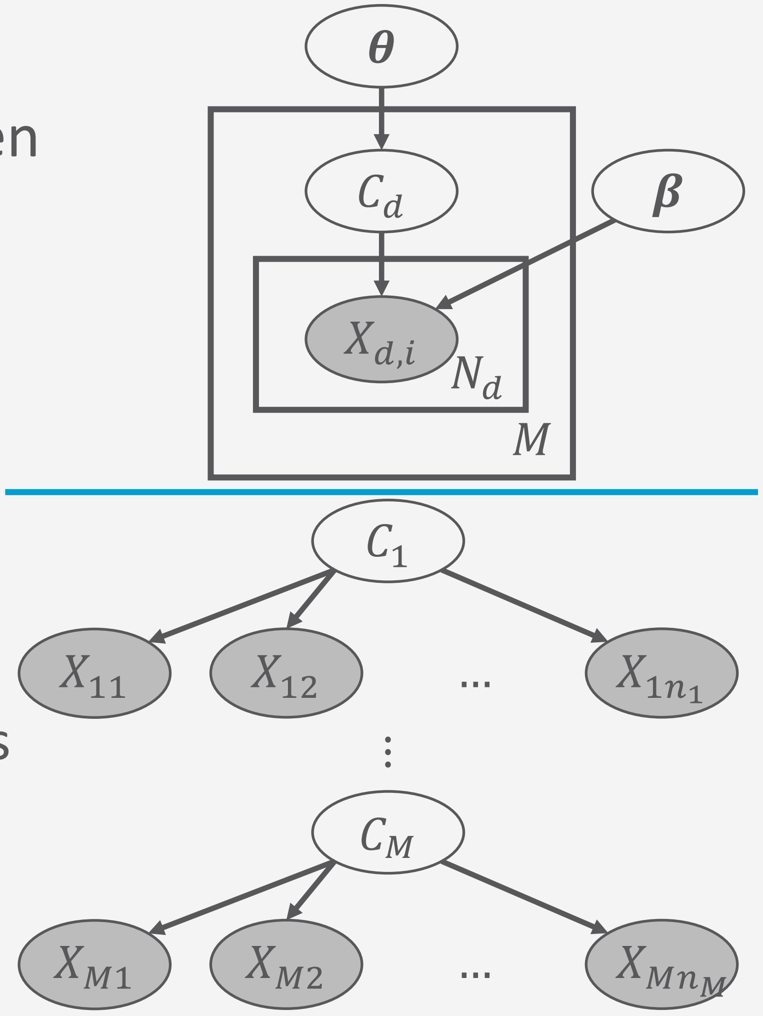
- Wiederkehrenden Strukturen in einer *Plate* zusammenfassen
  - Topic Modelle: Jedes Dokument ein Topic, Worte pro Dokument
- Plate: Box aus Subgraph + repräsentierte Anzahl I
  - Anstatt I teilweise Repräsentant i (im PGM Buch) oder repräsentierte Menge $\boldsymbol{I}$ angegeben
  - Repräsentiert ein propositionales BN:
    Subgraph (inkl. CPDs, ein/ausgehende Kanten) wird I mal instanziiert
    - Bzw. für jedes $i \in \boldsymbol{I}$
- Beispiel: Mixture of Unigrams
  - M Dokumente, N_d Worte im Dokument d
  - Plate-Modell oben rechts, Instanziierung unten rechts



David M. Blei, Andrew Y. Ng, and Michael I. Jordan: *Latent Dirichlet Allocation*. Journal of Machine Learning Research, 2003.

Plate Notation: Explizite Parameterdarstellung

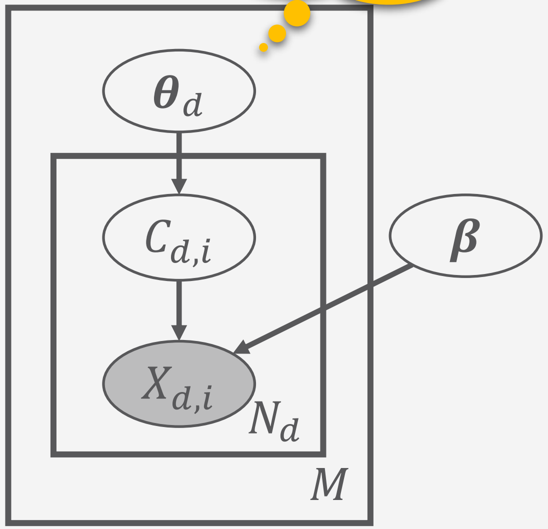
- Manchmal werden (Multinomial-) Parameter der Verteilungen explizit in das Plate Modell eingetragen
- Beispiel: Mixture of Unigrams
  - M Dokumente, N_d Worte im Dokument d
  - K mögliche Topics
  - Multinomial-Parameter $\boldsymbol{\theta} \in [0,1]^K$ für die A-priori-Verteilung $P(C)$ über $k \in \mathrm{Val}(C)$: $C \sim Mul(\boldsymbol{\theta})$
    - Insgesamt: Vektor der Länge K
    - ✓ Jedes C_d hat die gleiche A-priori-Verteilung $P(C)$
  - Multinomial-Parameter $\boldsymbol{\beta} \in [0,1]^N$ für die CPD $P(X|C)$ über das Lexikon $\mathcal{D} = \mathrm{Val}(X)$ für jedes $k \in \mathrm{Val}(C)$: $X \sim Mul(\boldsymbol{\beta}[k])$
    - Insgesamt: Matrix der Größe $K \times N$
    - ✓ Jedes X_{di} hat die gleiche CPD $P(X_i|C)$



David M. Blei, Andrew Y. Ng, and Michael I. Jordan: *Latent Dirichlet Allocation*. Journal of Machine Learning Research, 2003.

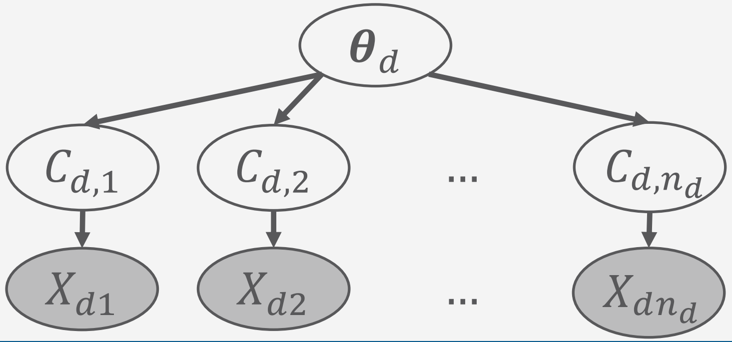
Topic Modellierung: Mixture of Topics

Warum reicht das so noch nicht?



- Andere Annahme: Dokumente bestehen aus einem Mix von Topics
- Pro Dokument $d \in \mathcal{C}$: Zufallsvariable X_i bezeichnet, welches Wort w aus einem Lexikon $\mathcal{D}$ an Position i in d vorkommt
  - I.e., $\mathrm{Val}(X_i) = \{w\}_{w\in\mathcal{D}}$
    - Bzw. $\mathrm{Val}(X_i) = \mathcal{D}$ mit $\mathcal{D}$ als Menge der Wörter im Lexikon
  - X_i unabhängig von X_j gegeben k:
  $$P(X_i, X_j|C) = P(X_i|C)P(X_j|C)$$
  - Annahme: Vorkommen von w unabhängig von Position:
  $$P(X_i = w) = P(X_j = w)$$
  - Soweit wie beim *Mixture of Unigrams*
- Änderungen:
  - Jedes Wort ergibt sich aus einem der Topics
  - Eigene A-priori-Topic-Verteilung $\boldsymbol{\theta}_d$ für jedes Dokument

Pro Dokument d:

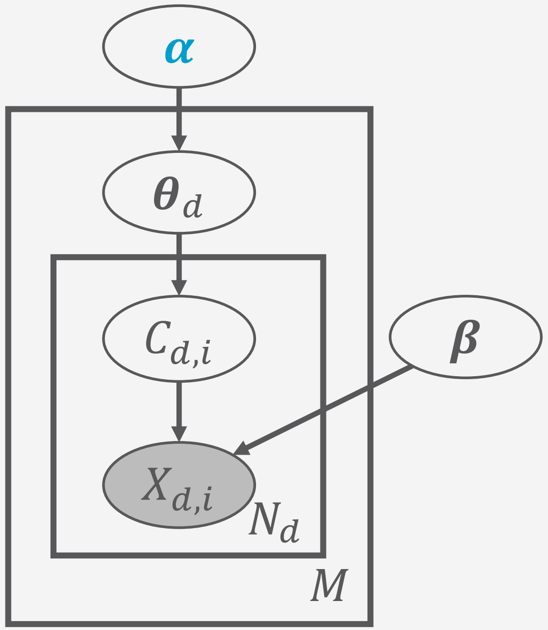


David M. Blei, Andrew Y. Ng, and Michael I. Jordan: *Latent Dirichlet Allocation*. Journal of Machine Learning Research, 2003.

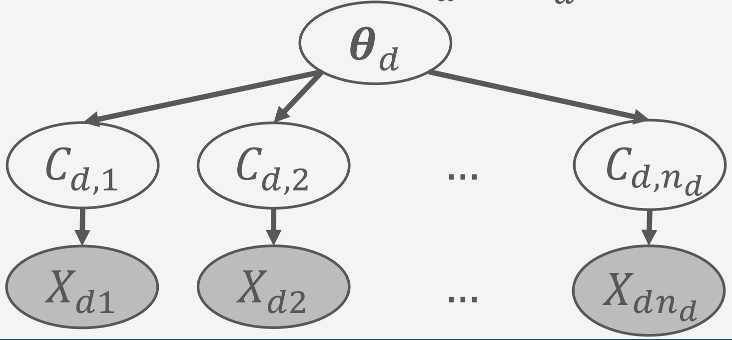
Topic Modellierung: Mixture of Topics

- Andere Annahme: Dokumente bestehen aus einem Mix von Topics
- Um eine eigene A-priori-Topic-Verteilung $\boldsymbol{\theta}_d$ für jedes Dokument darzustellen: Hyperparameter $\boldsymbol{\alpha}$
  - Als Elternknoten außerhalb der Plate: Ermöglicht $\boldsymbol{\theta}_d \neq \boldsymbol{\theta}_{d'}$
  - Gibt an, wie wahrscheinlich die unterschiedlichen Verteilungen $\boldsymbol{\theta}_d$ sind
    - *Verteilung über Verteilungen*
    - Erlaubt zu modellieren, dass es wahrscheinlicher ist, dass
      - Pro Dokument wenige Topics mit hoher Wahrscheinlichkeit auftreten

    oder
      - Pro Dokument jedes Topic mit etwa gleicher Wahrscheinlichkeit auftritt



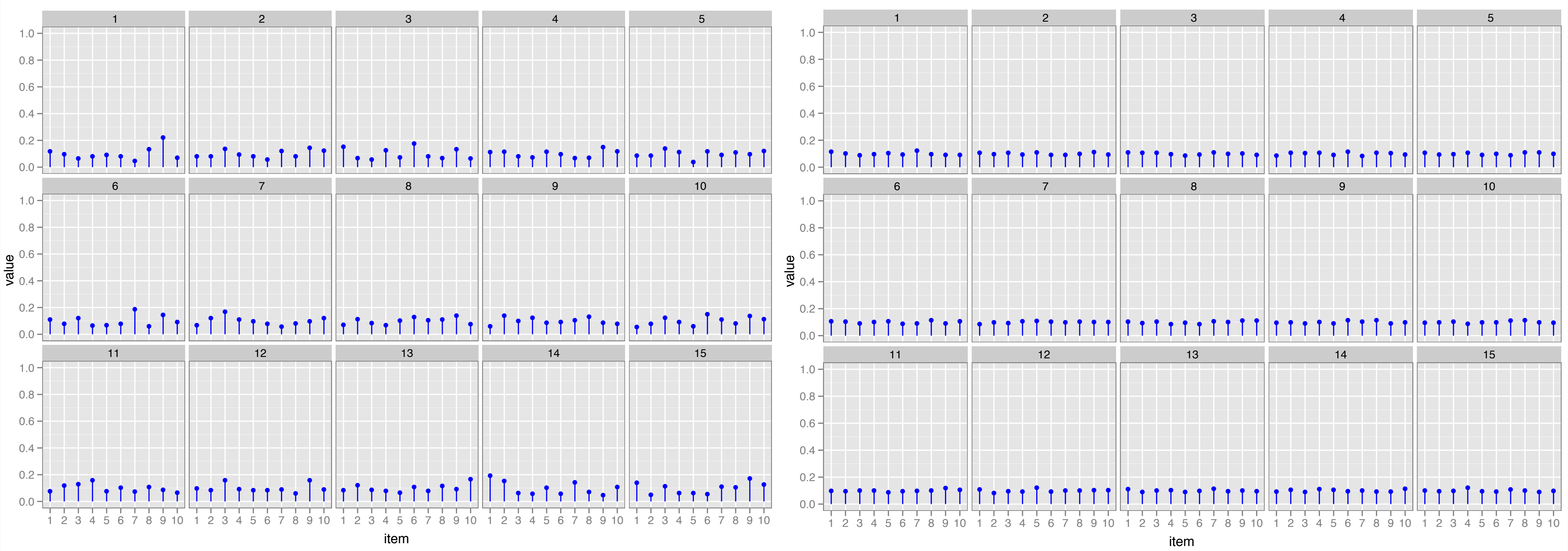
Pro Dokument d mit $\boldsymbol{\theta}_d \neq \boldsymbol{\theta}_{d'}$:



David M. Blei, Andrew Y. Ng, and Michael I. Jordan: *Latent Dirichlet Allocation*. Journal of Machine Learning Research, 2003.

Hyperparameter α

- Eher Gleichverteilung über $K = 10$ Topics

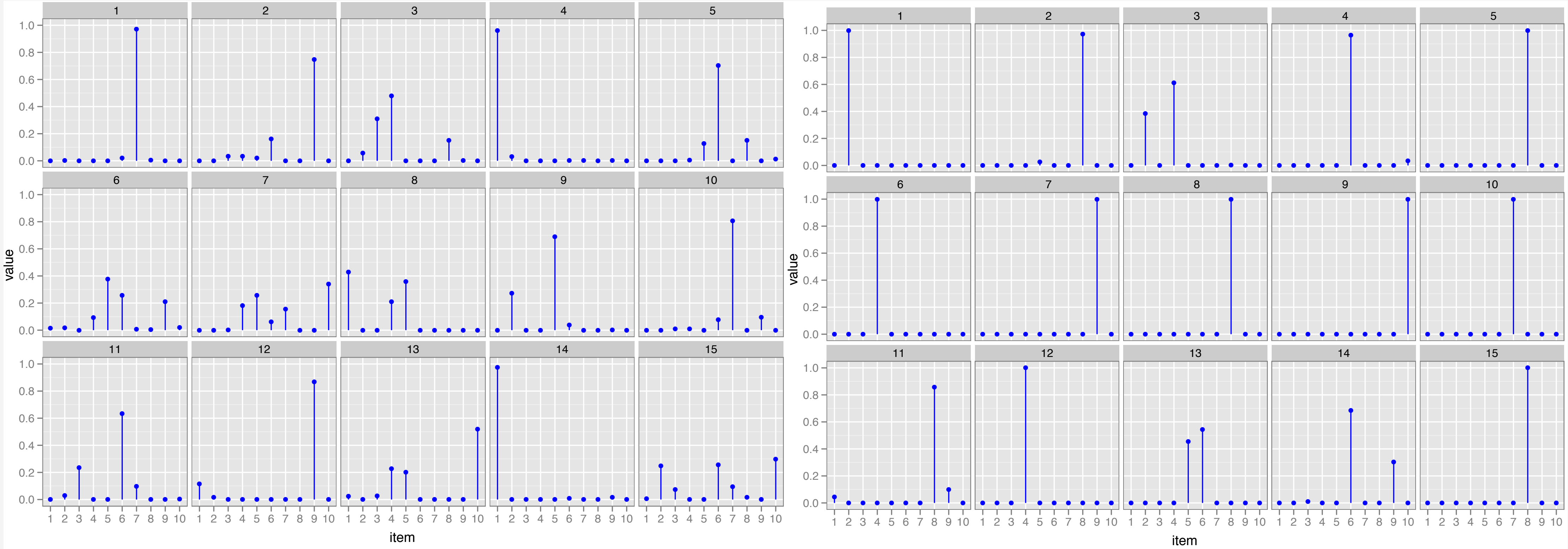


David M. Blei, Andrew Y. Ng, and Michael I. Jordan: *Latent Dirichlet Allocation*. Journal of Machine Learning Research, 2003.

Darstellung: David M. Blei.

Hyperparameter α

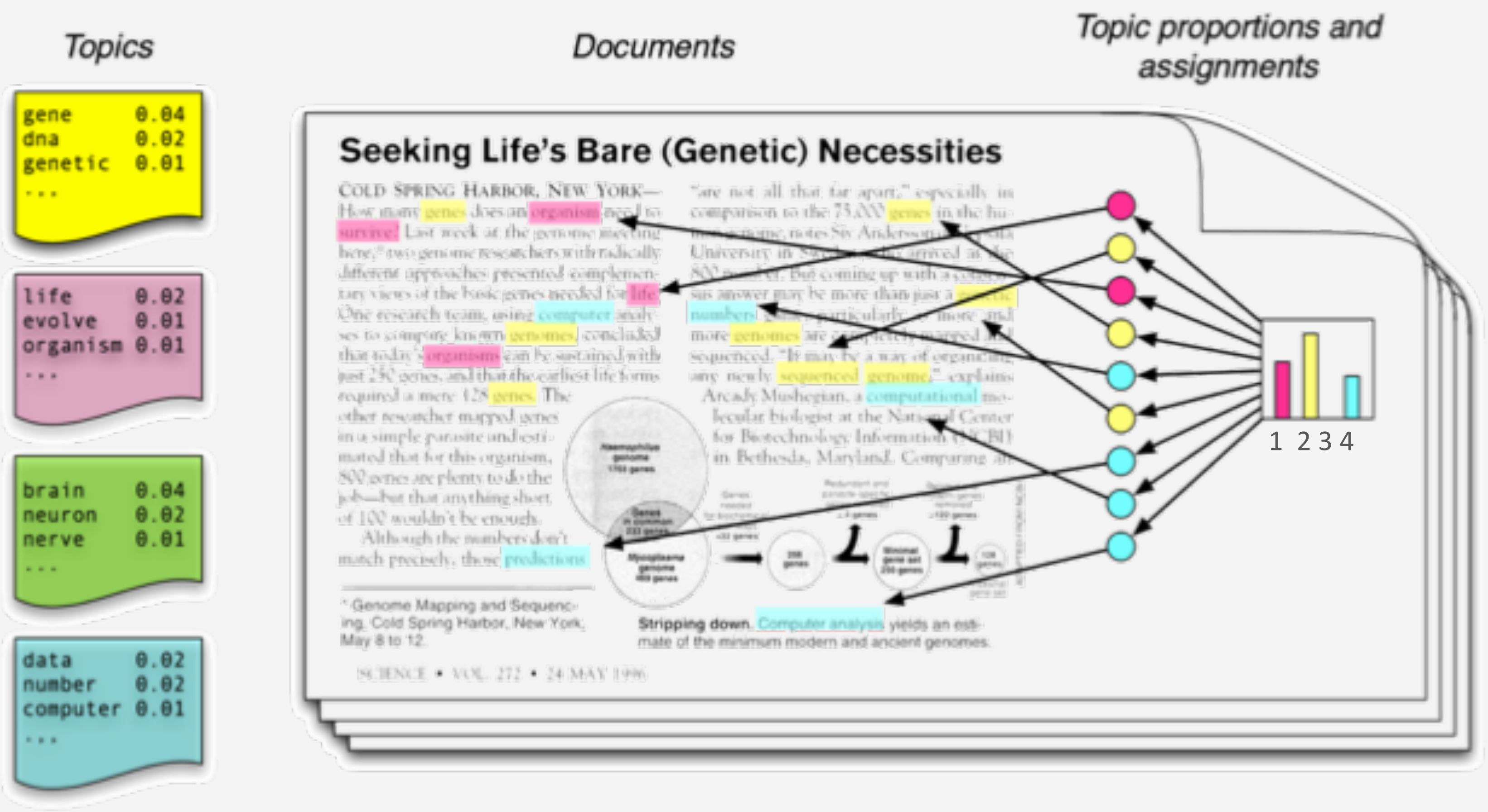
- Eher wenige Topics mit hoher Wahrscheinlichkeit

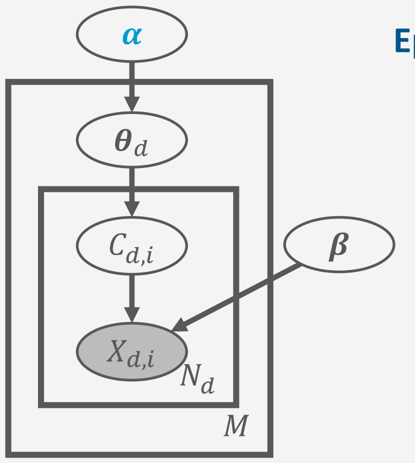


David M. Blei, Andrew Y. Ng, and Michael I. Jordan: *Latent Dirichlet Allocation*. Journal of Machine Learning Research, 2003.

Darstellung: David M. Blei.

Was sagt das Topic Modell jetzt aus?





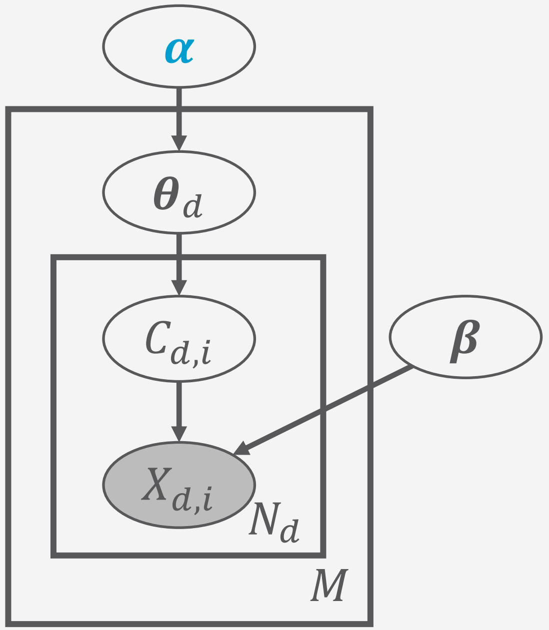
- Dokument hat eigene Verteilung über K Topics
- Topic repräsentiert als Verteilung über Wörter
  - Mehr und weniger wahrscheinliche Worte
  - Charakterisierbar durch top-l wahrscheinlichsten Worte

David M. Blei, Andrew Y. Ng, and Michael I. Jordan: *Latent Dirichlet Allocation*. Journal of Machine Learning Research, 2003.

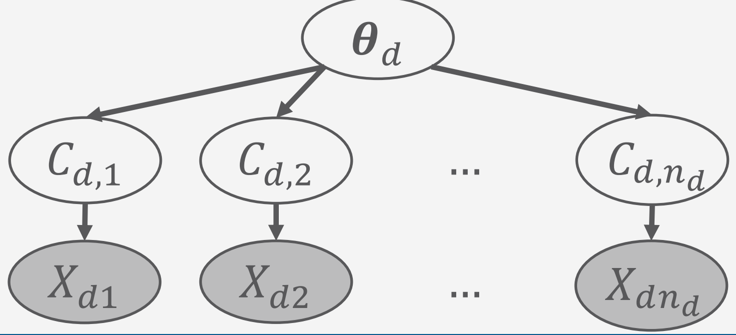
Darstellung: David M. Blei.

Inferenzaufgaben rund um Topic Modelle

- Gegeben ein Topic Modell für ein Korpus
  - Anfrage: Gegeben ein unbekanntes Dokument, welches Topic / welche Topic-Verteilung hat es?
    - Exakte + approximative Anfragenbeantwortung in PGMs: Thema 3 + 4
  - Generierung eines neuen Dokuments: Gegeben ein gewähltes Topic, generiere die Worte eines neuen Dokuments
    - Sampling aus PGMs: Thema 4
- Gegeben ein Korpus
  - Lerne ein Topic Modell: Thema 5
    - Gegeben: Worte in den Dokumenten (nach Vorverarbeitung), Anzahl Topics K, Hyperparameter $\boldsymbol{\alpha}$
    - Lerne: Topicverteilungen pro Dokument, Wortverteilungen pro Topic (Basis-Verfahren: *Latent Dirichlet Allocation, LDA*)



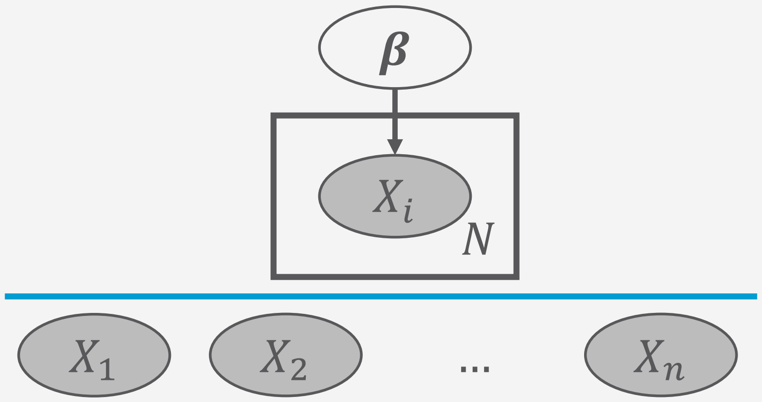
Pro Dokument d mit $\boldsymbol{\theta}_d \neq \boldsymbol{\theta}_{d'}$:



David M. Blei, Andrew Y. Ng, and Michael I. Jordan: *Latent Dirichlet Allocation*. Journal of Machine Learning Research, 2003.

Der Vollständigkeit halber: Dokumentenmodellierung ohne Topics

- Unigram: A-priori-Verteilung über Worte in Lexikon
  - X_i gibt an, ob ein Wort w_i aus Lexikon $\mathcal{D}$ vorkommt
  - Nur Bag-of-Words Annahme; *keine* Topics vorgesehen



fifth, an, of, futures, the, an, incorporated, a, a, the, inflation, most, dollars, quarter, in, is, mass

thrift, did, eighty, said, hard, 'm, july, bullish

that, or, limited, the

Automatisch generierte Sätze aus einem Unigramm-Modell ohne Vorverarbeitung

David M. Blei, Andrew Y. Ng, and Michael I. Jordan: *Latent Dirichlet Allocation*. Journal of Machine Learning Research, 2003.

Vor- und Nachteile einer BN-Kodierung

- Vorteil
  - Weniger Einträge als in vollständiger gemeinsamer Wahrscheinlichkeitsverteilung P_B
  - (Bedingte) Unabhängigkeiten, erforderliche CPDs direkt ersichtlich
  - Faktorisierung für Anfragebeantwortung nutzen
    - Thema 3: Exakte Inferenz in episodische PGMs

- Nachteil
  - BNs sind immer azyklisch → Schränkt die Ausdrucksstärke ein
    - Gegenseitige Einflüsse bzw. allgemeine Korrelationen nicht intuitiv darstellbar
- Alternative: Ungerichtete Modelle
  - Aufgabe der expliziten Repräsentation einer Richtung der Abhängigkeiten
    - Kann immer noch implizit vorhanden sein

Zwischenzusammenfassung

- Faktorisierung: Zerlegung einer vollständigen gemeinsamen Wahrscheinlichkeitsverteilung
  - Unter Ausnutzung von (bedingten) Unabhängigkeiten
- Bayes Netze
  - Repräsentation einer Faktorisierung auf Basis von (bedingten) Unabhängigkeiten
  - Syntax: gerichteter, azyklischer Graph mit CPDs $P\big(R \mid \mathrm{Pa}(R)\big)$ pro Knoten
  - Semantik:
    - Repräsentiert P_B als Produkt der CPDs: $P_B = P(\boldsymbol{R}) = \prod_{R \in \boldsymbol{R}} P\big(R \mid \mathrm{Pa}(R)\big)$
    - Repräsentiert (bedingte) Unabhängigkeiten: $R \perp \mathrm{Ndesc}(R) \mid \mathrm{Pa}(R)$
  - Anwendung
    - Naive Bayes Klassifizierer
      - Topic Modellierung: Mixture of Unigrams, Mixture of Topics

Überblick: 2. Episodische PGMs

A. *Probabilistische Modellierung*
   - Zufallsvariablen, vollständige gemeinsame Wahrscheinlichkeitsverteilung, Speicherkomplexität
   - Inferenzaufgaben, Komplexität

B. *Gerichtete Modelle: Bayes Netze (BNs)*
   - (Bedingte) Unabhängigkeiten, Faktorisierung
   - Syntax, Semantik, graphische Darstellung, Speicherkomplexität

C. ***Ungerichtete Modelle: Faktormodelle***
   - Syntax, Semantik, graphische Darstellungen und deren Unterschiede, Speicherkomplexität
   - Umwandlung von BNs zu Faktormodellen und deren graphischen Darstellungen

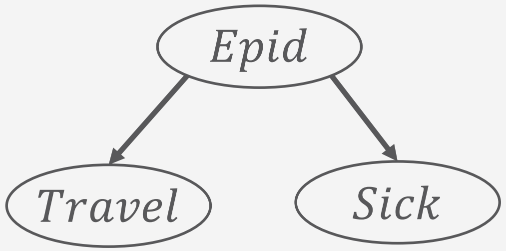
D. *Unabhängigkeiten in PGMs*
   - Lokale, globale und paarweise Unabhängigkeiten
   - Äquivalenzbedingungen von Faktormodellen und BNs

Ungerichtete Modelle

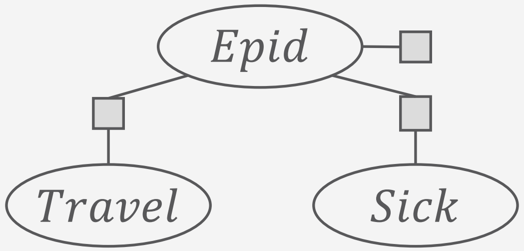
- Faktorisierung einer vollständigen gemeinsamen Wahrscheinlichkeitsverteilung P_R durch eine Menge von Faktoren
  - Manchmal auch *Potentialfunktionen* genannt
- Zwei graphische Darstellungen solcher Faktormodelle:
  - Faktorgraphen (FG): explizite Faktoren
  - Markov Netze (MNs): implizite Faktoren über die Cliquen im Graph

$Epid$	$Travel$	$Sick$	P
$false$	$false$	$false$	0.51
$false$	$false$	$true$	0.09
$false$	$true$	$false$	0.17
$false$	$true$	$true$	0.03
$true$	$false$	$false$	0.09
$true$	$false$	$true$	0.03
$true$	$true$	$false$	0.06
$true$	$true$	$true$	0.02

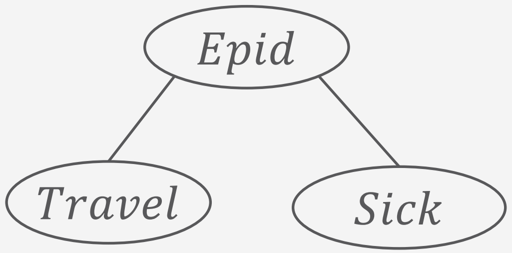
Vollständige gemeinsame Wahrscheinlichkeitsverteilung



Bayes Netzwerk (BN)



Faktorgraph (FG)



Markov Netzwerk (MN)

Faktoren

- Faktor $f = \phi(R_1, \ldots, R_k)$
  - Argumente $R_1, \ldots, R_k$: Zufallsvariablen
    - $\mathrm{rv}(f_i) = \{R_1, \ldots, R_k\}$
  - Potentialfunktion ϕ: Reell-wertige, positive Funktion
    $$\phi\colon \times_{i=1}^{k} \mathrm{Val}(R_i) \longrightarrow \mathbb{R}^{0,+}$$
    - Bildet auf Potentiale ab
      - Muss keine Wahrscheinlichkeitsverteilung sein (allg. Verteilung)
      - Misst so etwas wie Korrelation, Kompatibilität, Affinität, Gewichtung, ...
      - Herkunft aus der Signalverarbeitung: Potential in elektrischen Feldern
    - Mindestens ein Potential > 0
      - Muss gelten, damit nachher die gemeinsame vollständige Wahrscheinlichkeitsverteilung nicht 0 wird
    - Notieren als Tabelle, Liste, ...
  - Beispiel: Faktor $f = \phi(Epid, Travel, Sick)$
    - Abbildung gemäß Tabelle rechts

$Epid$	$Travel$	$Sick$	ϕ
$false$	$false$	$false$	20
$false$	$false$	$true$	24
$false$	$true$	$false$	28
$false$	$true$	$true$	8
$true$	$false$	$false$	5
$true$	$false$	$true$	6
$true$	$true$	$false$	7
$true$	$true$	$true$	2

Faktormodell

- Gegeben eine Menge von Zufallsvariablen $\boldsymbol{R} = \{R_1, \dots, R_n\}$
- Menge von Faktoren $F = \{f_i\}_{i=1}^{n'}$
  = Faktormodell
  - Faktoren f_i mit Argumenten $R_1, \dots, R_k \in \boldsymbol{R}$
    - Bzw. $\boldsymbol{R} := \bigcup_{i=1}^{n'} \mathrm{rv}(f_i)$
  - Beispiel: $F = \{f_0, f_1, f_2\}$
    - $\boldsymbol{R} = \{Epid, Travel, Sick\}$
    - $f_0 = \phi_0(Epid)$
    - $f_1 = \phi_1(Epid, Travel)$
    - $f_2 = \phi_2(Epid, Sick)$
      - Faktoren sind die CPDs des BN-Beispiels

$Epid$	ϕ_0
$false$	0.8
$true$	0.2

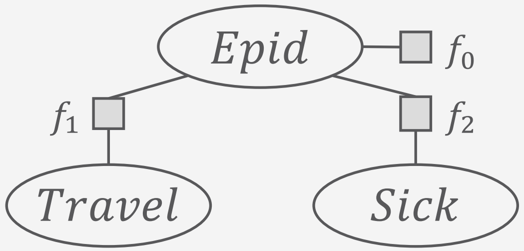
$Epid$	$Travel$	ϕ_1
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

$Epid$	$Sick$	ϕ_2
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

Faktormodell: Eine erste graphische Darstellung

- Graphische Darstellung eines Faktormodell $F = \{f_i\}_{i=1}^{n'}$ als Faktorgraph (FG)
  - Für jedes $R \in \boldsymbol{R}$: Variablenknoten in FG (Ellipse)
  - Für jedes $f \in F$: Faktorknoten in FG (Box)
  - Für jedes Argument $R \in \mathrm{rv}(f), f \in F$: Kante zwischen Variablenknoten für R und Faktorknoten für f
  - D.h. ein FG G ist ein bipartiter Graph (V, E) mit $V = \boldsymbol{R} \cup F$
  - Beispiel: $F = \{f_0, f_1, f_2\}$ und $\boldsymbol{R} = \{Epid, Travel, Sick\}$
    - Variablenknoten für $Epid, Travel, Sick$
    - Faktorknoten für f_0, f_1, f_2
    - Kanten zwischen f_1 und $Epid, Travel$; zwischen f_2 und $Epid, Sick$; zwischen f_o und $Epid$
      - Beispiel ist eine FG-Darstellung des Beispiel-BNs

$Epid$	ϕ_0
$false$	0.8
$true$	0.2



$Epid$	$Travel$	ϕ_1
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

$Epid$	$Sick$	ϕ_2
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

Faktormodell: Semantik

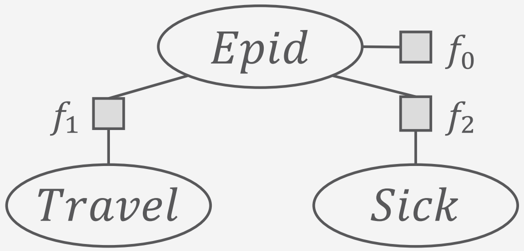
- Semantik eines Faktormodell $F = \{f_i\}_{i=1}^{n'}$
  - Vollständige gemeinsame Wahrscheinlichkeitsverteilung P_F als normalisiertes Produkt der Faktoren:

$$P_F = \frac{1}{Z} \prod_{i=1}^{n'} \phi_i(R_1, \dots, R_k)$$

$$Z = \sum_{r_1 \in \mathrm{Val}(R_1)} \dots \sum_{r_n \in \mathrm{Val}(R_n)} \prod_{i=1}^{n'} \phi_i(r_1, \dots, r_k)$$

  - Multiplikation von Faktoren funktioniert wie bisher
  - Normalisierung notwendig, weil die f_i keine Wahrscheinlichkeitsverteilungen sein müssen
    - $Z = P(.)$ die leere Anfrage
  - P_F auch Gibbs Verteilung genannt

$Epid$	ϕ_0
$false$	0.8
$true$	0.2



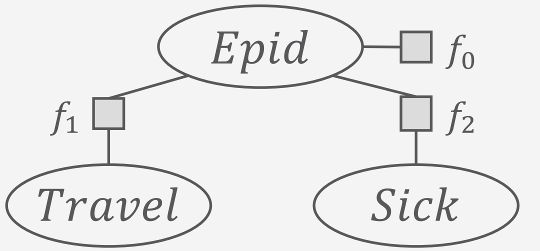
$Epid$	$Travel$	ϕ_1
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

$Epid$	$Sick$	ϕ_2
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

Faktormodelle: Beispiele

- Faktorgraph mit $F = \{f_0, f_1, f_2\}$ und $\boldsymbol{R} = \{Epid, Travel, Sick\}$:
  - $P_F = \frac{1}{Z} f_1 \cdot f_2 \cdot f_3 = P_B$
  - Weil Faktoren CPDs sind: $Z = 1$
    - Macht den Vorteil eines BNs noch einmal deutlich: Normalisierungskonstante Z ist 1, muss also nicht durch eine leere Anfrage berechnet werden

$Epid$	ϕ_0
$false$	0.8
$true$	0.2



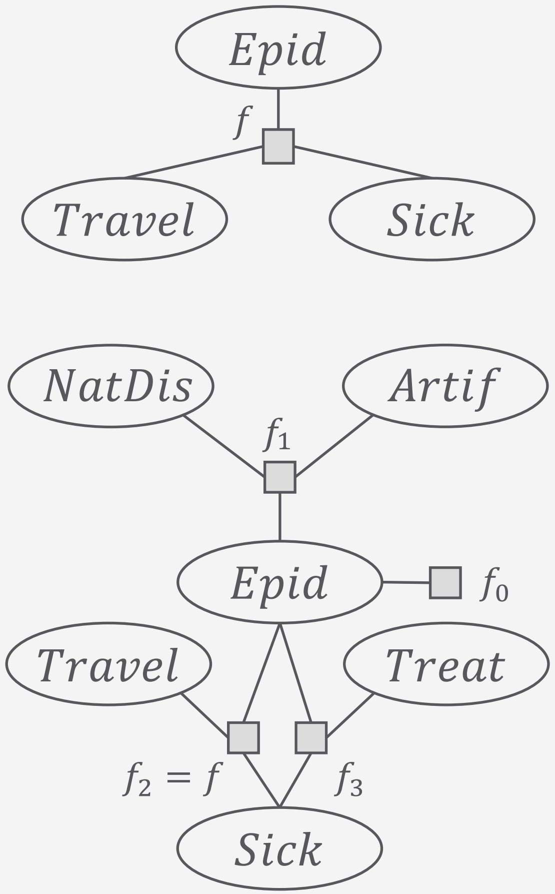
$Epid$	$Travel$	$Sick$	P_F
$false$	$false$	$false$	0.51
$false$	$false$	$true$	0.09
$false$	$true$	$false$	0.17
$false$	$true$	$true$	0.03
$true$	$false$	$false$	0.09
$true$	$false$	$true$	0.03
$true$	$true$	$false$	0.06
$true$	$true$	$true$	0.02

$Epid$	$Travel$	ϕ_1
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

$Epid$	$Sick$	ϕ_2
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

Faktormodelle: Beispiele

- Faktor von der Faktordefinition
  - $Z = 100$ als Summe der Potentiale
    - Ergibt die Beispielverteilung vom Anfang, die keine bedingten Unabhängigkeiten aufwies
- Könnte Teil eines größeren Modells sein: Faktormodell $F = \{f_i\}_{i=0}^{3}$
  - $f_0 = \phi_0(Epid)$ als Quasi-Apriori-Verteilung
  - $f_1 = \phi_1(Epid, NatDis, Artif)$ als Quasi-CPD
  - $f_2 = \phi_2(Epid, Travel, Sick) = f$
  - $f_3 = \phi_3(Epid, Sick, Treat)$ ähnlich zu f
  - Repräsentiert $P_F = \frac{1}{Z}\prod_{i=1}^{3} f_i$



$Epid$	$Travel$	$Sick$	ϕ
$false$	$false$	$false$	20
$false$	$false$	$true$	24
$false$	$true$	$false$	28
$false$	$true$	$true$	8
$true$	$false$	$false$	5
$true$	$false$	$true$	6
$true$	$true$	$false$	7
$true$	$true$	$true$	2

ϕ
0.20
0.24
0.28
0.08
0.05
0.06
0.07
0.02

Faktormodell: Speicherkomplexität

- Gegeben ein Faktormodell $F = \{f_i\}_{i=1}^{n'}$ über $\boldsymbol{R}$
- Speicherkomplexität: $O(n' \cdot r^m)$
  - $m = \max_{f \in F} |\mathrm{Nb}(f)| = \max_{f \in F} |\mathrm{rv}(f)|$
  - $r = \max_{R \in \boldsymbol{R}} |\mathrm{Val}(R)|$
  - Herleitung

$$\sum_{f \in F} \prod_{R \in \mathrm{rv}(f)} |\mathrm{Val}(R)| \leq \sum_{f \in F} \prod_{R \in \mathrm{rv}(f)} \max_{R \in \boldsymbol{R}} |\mathrm{Val}(R)| = \sum_{f \in F} \prod_{R \in \mathrm{rv}(f)} r = \sum_{f \in F} r^{|\mathrm{rv}(f)|} \leq \sum_{f \in F} \overbrace{r^{\max_{f \in F} |\mathrm{rv}(f)|}}^{\text{Größtmöglicher Faktor}} = \sum_{f \in F} r^m = n' \cdot r^m$$

  - Nicht mehr exponentiell in $n = |\boldsymbol{R}|$, sondern in Anzahl Nachbarknoten m
    - Annahme: $m \ll n$ und n' nicht exponentiell abhängig von n, so dass $O(n' \cdot r^m) \ll O(r^n)$
      - Realistische Szenarien: $n' < n$ oder in gleicher Größenordnung
  - Vergleich BN: $O(n \cdot r^m)$

Umwandlung von BNs in Faktormodelle

Basierend auf den CPDs im BN

- Gegeben ein BN B über $\boldsymbol{R} = \{R_1, \dots, R_n\}$ mit vollständiger gemeinsamer Verteilung P_B
- Leeres Faktormodell $F = \emptyset$
- Für jedes $R \in \boldsymbol{R}$
  - Wandle die CPD $P\big(R \mid \mathrm{Pa}(R)\big)$ in einen Faktor $\phi\big(R, \mathrm{Pa}(R)\big)$ um
    - ❖ I.e., $\phi(r, \boldsymbol{a}) = P(r|\boldsymbol{a})$ für alle $r \in \mathrm{Val}(R)$ und $\boldsymbol{a} \in \mathrm{Val}\big(\mathrm{Pa}(R)\big)$
  - Füge $\phi\big(R, \mathrm{Pa}(R)\big)$ zu F hinzu
- Gebe F zurück

❖ Es gilt: $P_B = P_F$ mit $Z = 1$ in P_F

❖ Optional: Nach Umwandlung von B in F, baue ein FG aus F

Umwandlung von BNs in FGs

Basierend auf der Graphstruktur

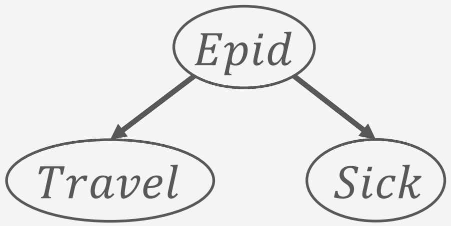
- Gegeben ein BN $B = (V_B, E_B)$ über $\boldsymbol{R} = \{R_1, \dots, R_n\}$

Für ein FG $G = (V_G, E_G)$

- $V_G := V_B$ (bzw. $\boldsymbol{R}$; Variablenknoten)
- Für jedes $R \in \boldsymbol{R}$
  - Wandle $P\big(R \mid \mathrm{Pa}(R)\big)$ in Faktor $f = \phi\big(R, \mathrm{Pa}(R)\big)$ um (wie auf vorheriger Folie)
  - $V_G \leftarrow V_G \cup \{f\}$ (Faktorknoten)
  - $E_G \leftarrow E_G \cup \big\{\{R, f\}\big\}$ (Kante von R zu f; einzige Kante, wenn $\mathrm{Pa}(R) = \emptyset$)
  - Für jedes $R' \in \mathrm{Pa}(R)$
    - $E_G \leftarrow E_G \cup \big\{\{R', f\}\big\}$ (Kante von Elternknoten zu f)
- Gebe G zurück

Faktormodelle: Beispiel Umwandlung BN in Faktormodell bzw. FG

- BN $B = (V_B, E_B)$
  - $V_B = \{Epid, Travel, Sick\}$
  - $E_B = \{(Epid, Travel), (Epid, Sick)\}$
  - CPDs $P(Epid)$, $P(Travel|Epid)$, $P(Sick|Epid)$
- Faktormodell $F = \{f_0, f_1, f_2\}$
  - $f_0 = \phi(Epid)$
  - $f_1 = \phi(Epid, Travel)$
  - $f_2 = \phi(Epid, Sick)$
- FG $G = (V_G, E_G)$
  - $V_G = V_B \cup \{f_0, f_1, f_2\}$
  - $E_G = \{\{Epid, f_0\}, \{Travel, f_1\}, \{Epid, f_1\}, \{Sick, f_2\}, \{Epid, f_2\}\}$

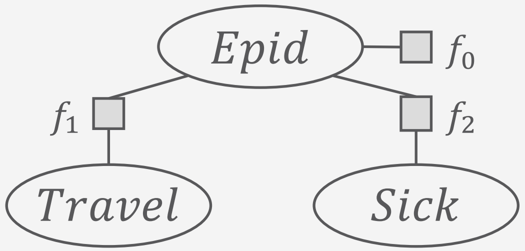


$Epid$	P
$false$	0.8
$true$	0.2

$Epid$	$Travel$	P
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

$Epid$	$Sick$	P
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

$Epid$	ϕ_0
$false$	0.8
$true$	0.2

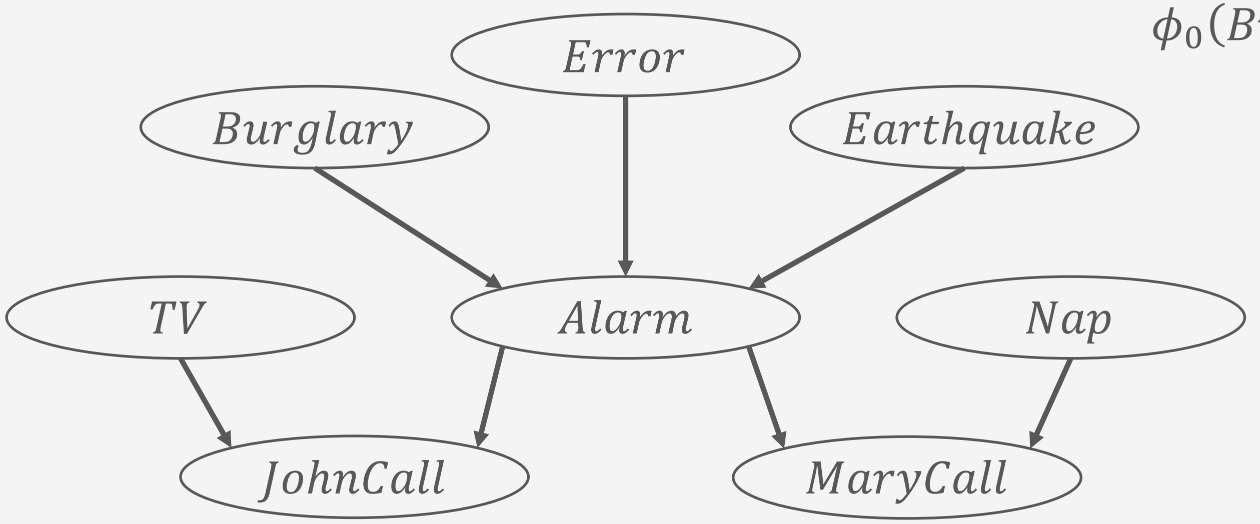


$Epid$	$Travel$	ϕ_1
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

$Epid$	$Sick$	ϕ_2
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

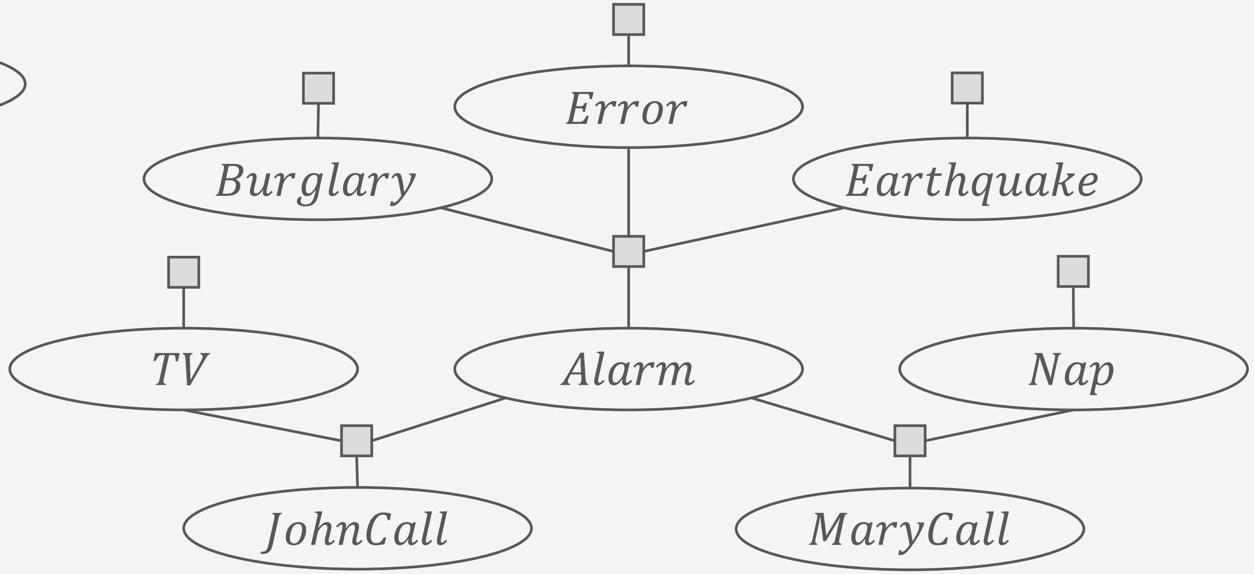
Faktormodelle: Beispiel Umwandlung BN in Faktormodell bzw. FG

- Ein Beispiel mit mehreren Elternknoten



$$P(Burglary), P(Error), P(Earthquake), P(TV), P(Nap)$$
$$P(Alarm|Burglary, Error, Earthquake)$$
$$P(JohnCall \mid TV, Alarm)$$
$$P(MaryCall \mid Nap, Alarm)$$

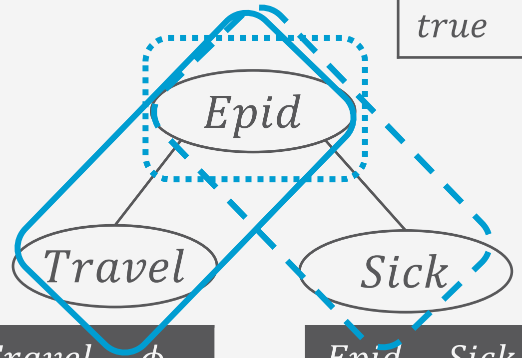
$$\phi_0(Burglary), \phi_1(Error), \phi_2(Earthquake), \phi_3(TV), \phi_4(Nap)$$
$$\phi_5(Alarm, Burglary, Error, Earthquake)$$
$$\phi_6(JohnCall, TV, Alarm)$$
$$\phi_7(MaryCall, Nap, Alarm)$$



Faktormodelle: Eine zweite graphische Darstellung

- Graphische Darstellung eines Faktormodell $F = \{f_i\}_{i=1}^{n'}$ als Markov Netz (MN)
  - Auch *Markov Random Field (MRF)* genannt
- Für jedes $R \in \boldsymbol{R}$: Knoten in MN (Ellipse)
- Für jedes $f \in F$: Paarweise Kanten zwischen allen Knoten $\mathrm{rv}(f)$
  - Vollständiger Subgraph pro f bzw. $\mathrm{rv}(f)$ ist Clique in MN
    - Zu jeder Clique gehört das entsprechende f, deshalb Faktoren auch Cliquenpotentiale genannt (*clique potentials*)
    - Faktoren erscheinen nicht explizit im Graph
- D.h. ein MN ist ein Graph (V, E) mit $V = \boldsymbol{R}$
- Beispiel: $F = \{f_0, f_1, f_2\}$ und $\boldsymbol{R} = \{Epid, Travel, Sick\}$
  - Knoten für $Epid, Travel, Sick$
  - Kanten zwischen $Epid, Travel$ wegen f_1; zwischen $Epid, Sick$ wegen f_2; keine weitere Kante wegen f_0
  - MN des Beispiel-BNs

$Epid$	ϕ_0
$false$	0.8
$true$	0.2

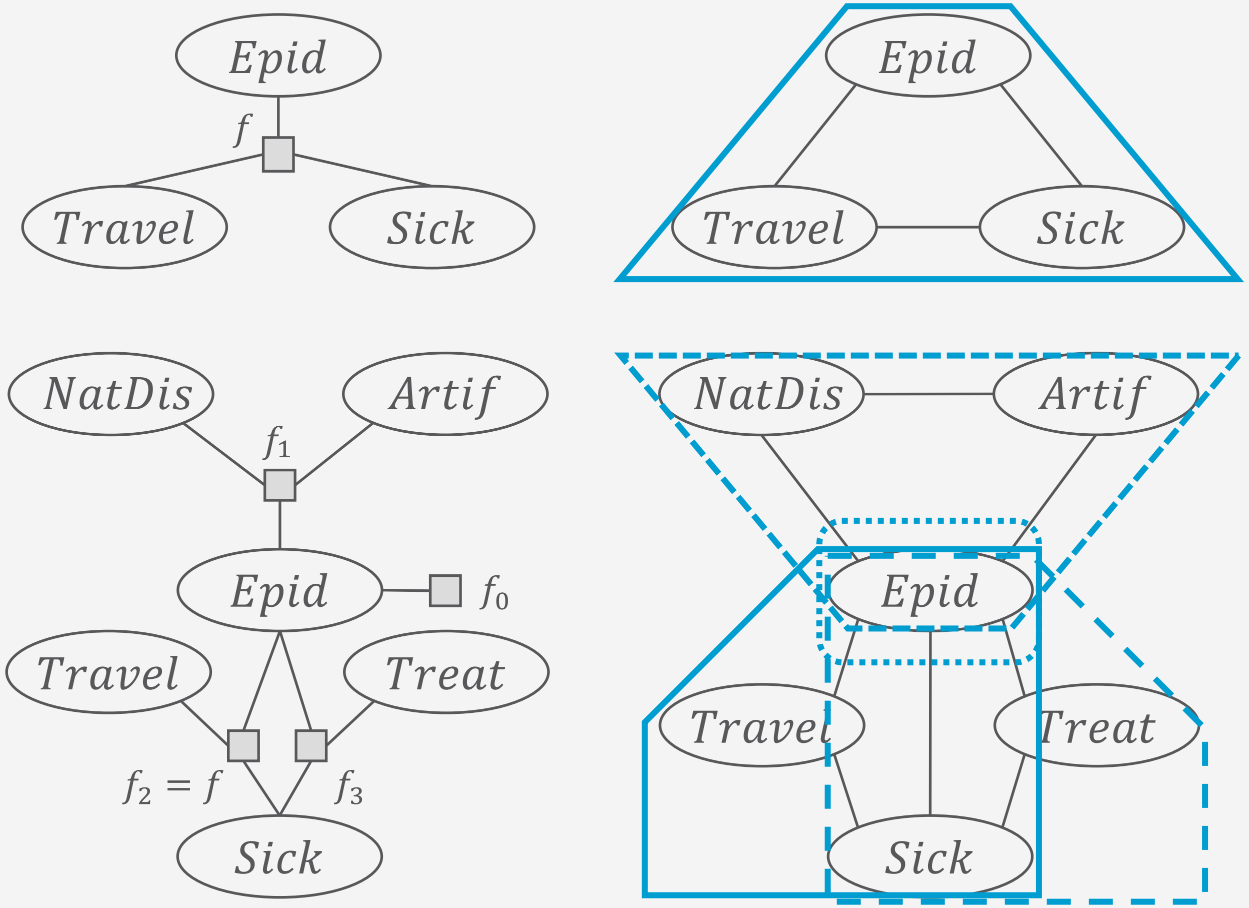


$Epid$	$Travel$	ϕ_1
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

$Epid$	$Sick$	ϕ_2
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

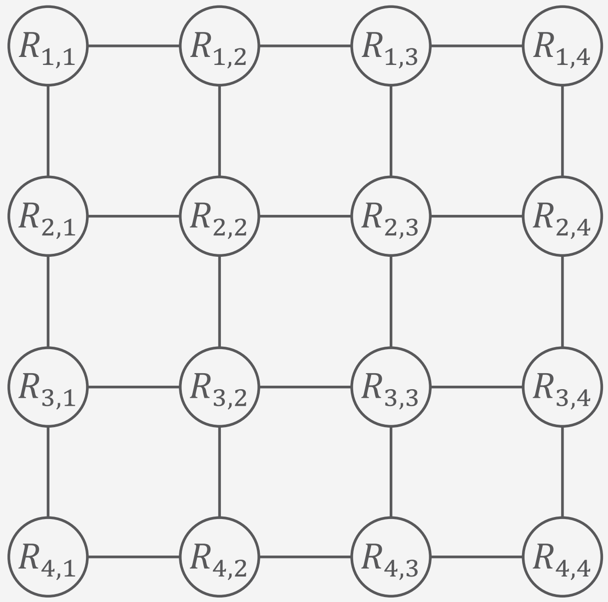
Faktormodelle: Beispiele als Markov Netzwerke

- Faktor von der Faktordefinition
  - $Z = 100$ als Summe der Potentiale
    - Ergibt die Beispielverteilung vom Anfang, die keine bedingte Unabhängigkeiten aufwies
- Könnte Teil eines größeren Modells sein: Faktormodell $F = \{f_i\}_{i=0}^{3}$
  - $f_0 = \phi_0(Epid)$ als Quasi-Apriori-Verteilung
  - $f_1 = \phi_1(Epid, NatDis, Artif)$ als Quasi-CPD
  - $f_2 = \phi_2(Epid, Travel, Sick) = f$
  - $f_3 = \phi_3(Epid, Sick, Treat)$ ähnlich zu f
  - Repräsentiert $P_F = \frac{1}{Z}\prod_{i=0}^{3} f_i$

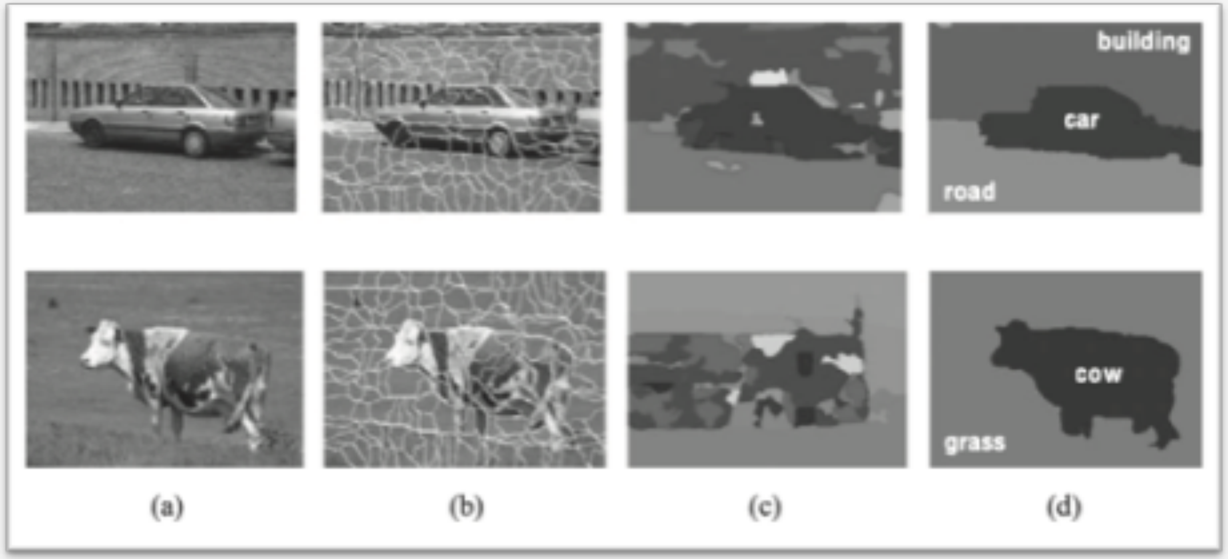


Einsatz von Markov Netzen

- Paarweise MNs (im Gitter)
  - Faktoren pro Knoten: $\phi(R)$ (Einercliquen; Knotenpotentiale)
  - Faktoren pro Kante: $\phi(R, R')$ (Zweiercliquen; Kantenpotentiale)
- U.a. im Bereich Computer Vision eingesetzt
  - Netzwerk: Gitter = Bild
    - Zufallsvariablen = Knoten = *Pixel*
    - Faktoren = Kanten = *Interaktion zwischen benachbarten Pixeln*
    - Potentiale als Strafe für Diskrepanz interpretiert (niedriger Wert besser)
  - Aufgabe: Rauschunterdrückung
    - Ziel: „Wahren“ Pixel-Wert wieder herstellen
      - Knotenpotential: Bestrafung einer (hohen) Diskrepanz zwischen inferiertem und beobachtetem Pixel-Wert
      - Kantenpotentiale: Bestrafung von Diskrepanz zwischen benachbarten inferierten Pixel-Werten
        → Achtung: wahre Diskrepanzen bei Kanten zwischen Objekten oder Regionen



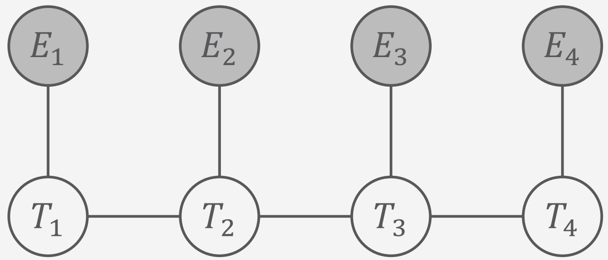
Einsatz von Markov Netzen



- Aufgabe: Bildsegmentierung
  - Knoten = Superpixel (kohärente Regionen)
    - Features (Farbe, Textur, Position) extrahieren, evtl. auf Label / Cluster abbilden (Dimensionsreduktion)
    - Knotenpotentiale als Funktion der Features
  - Kanten zwischen Superpixeln, die aneinander grenzen
    - Kantenpotentiale: Bestrafung von Diskrepanz zwischen benachbarten Superpixeln *gegeben Features*
    - Domänenwissen kodierbar: Tiger eher neben Vegetation als Wasser, Wasser unter Vegetation, Autos auf Straßen, Himmel über allem
  - Abbildung: (a) Original (b) Superpixel (c) Nur mit Knotenpotentialen segmentiert (d) Mit einem paarweisen MN segmentiert

Abbildung: PGM, S. 114

Einsatz von Markov Netzen



- MN für Bildsegmentierung eigentlich ein *conditional random field (CRF)*
- Kodierung einer bedingten Verteilung $P(\boldsymbol{T} \mid \boldsymbol{E})$ mit m Faktoren $\phi_i(\boldsymbol{D}_i)$
  - $\boldsymbol{T}$ Menge von Zielvariablen (***t**arget*)
  - $\boldsymbol{E}$ Menge von beobachteten Variablen (***e**vidence*); $\boldsymbol{E} \cap \boldsymbol{T} = \emptyset$
  - $\boldsymbol{D}_i \nsubseteq \boldsymbol{E} \rightarrow$ Keine Verteilungen nur über (Untermengen von) $\boldsymbol{E}$
    - Erlaubt komplexe (kontinuierliche) Features, deren Zusammenhang nur schwer zu kodieren oder nicht verstanden ist
  - Ergibt ungerichteten Graph (V, E) mit $V = \boldsymbol{T} \cup \boldsymbol{E}$, Ausgrauen der Variablen $\boldsymbol{E}$ (keine Knotenpotentiale)
  - Semantik:

$$P(\boldsymbol{T} \mid \boldsymbol{E}) = \frac{1}{Z(\boldsymbol{E})} \prod_{i=1}^{m} \phi_i(\boldsymbol{D}_i) \qquad Z(\boldsymbol{E}) = \sum_{\boldsymbol{t} \in \mathrm{Val}(\boldsymbol{T})} \tilde{P}(\boldsymbol{T} = \boldsymbol{t}, \boldsymbol{E})$$

  - Andere Normalisierung im Vergleich zur Semantik von Faktormodellen (MNs)
    - Unterschiedliche Werte für jede Zuweisung $\boldsymbol{e}$ an $\boldsymbol{E}$
  - Bildsegmentierung: $\boldsymbol{E}$ = Features der Superpixel, Zusammenhang von Farbe / Textur / Position komplex $\rightarrow$ Bedingte Verteilung über Segmentzuweisungen *gegeben der Features*

Umwandlung von BNs in MNs

Basierend auf der Graphstruktur

- Gegeben ein BN $B = (V_B, E_B)$ über $\boldsymbol{R} = \{R_1, \dots, R_n\}$

Für ein MN $M = (V_M, E_M)$

- $V_M := V_B$ (bzw. $\boldsymbol{R}$)
- Für jedes $R \in \boldsymbol{R}$
  - Für jedes $R' \in \mathrm{Pa}(R)$
    - $E_M \leftarrow E_M \cup \{\{R, R'\}\}$ (gerichtete Kante wird zu ungerichteter Kante)
    - Für jedes $R'' \in \mathrm{Pa}(R)$
      - $E_M \leftarrow E_M \cup \{\{R', R''\}\}$ (Elternknoten werden paarweise mit einer Kante verbunden)

❖ Nicht effizient aufgeschrieben, da einige Kanten doppelt hinzugefügt werden

❖ Fügt möglicherweise mehr Kanten hinzu als im BN sind, um die Clique zur CPD $P(R \mid \mathrm{Pa}(R))$ zu erzeugen

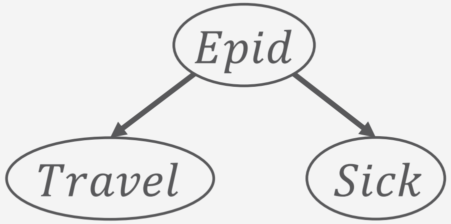
❖ Faktor zu einer Clique wie bisher aus der CPD $P(R \mid \mathrm{Pa}(R))$ erzeugen

Einen DAG moralisieren

- Gerichtete Kanten in ungerichtete Kanten umwandeln
- Elternknoten durch Kanten verbinden

Faktormodelle: Beispiel Umwandlung BN in MN

- BN $B = (V_B, E_B)$
  - $V_B = \{Epid, Travel, Sick\}$
  - $E_B = \{(Epid, Travel), (Epid, Sick)\}$
  - CPDs $P(Epid)$, $P(Travel|Epid)$, $P(Sick|Epid)$
- MN $M = (V_M, E_M)$
  - $V_M = V_B$
  - $E_M = \{\{Epid, Travel\}, \{Epid, Sick\}\}$
  - Mit den Faktoren
    - $f_0 = \phi(Epid)$
    - $f_1 = \phi(Epid, Travel)$
    - $f_2 = \phi(Epid, Sick)$

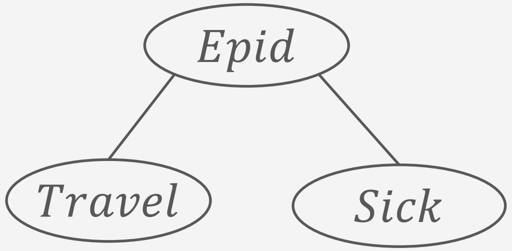


$Epid$	P
$false$	0.8
$true$	0.2

$Epid$	$Travel$	P
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

$Epid$	$Sick$	P
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

$Epid$	ϕ_0
$false$	0.8
$true$	0.2

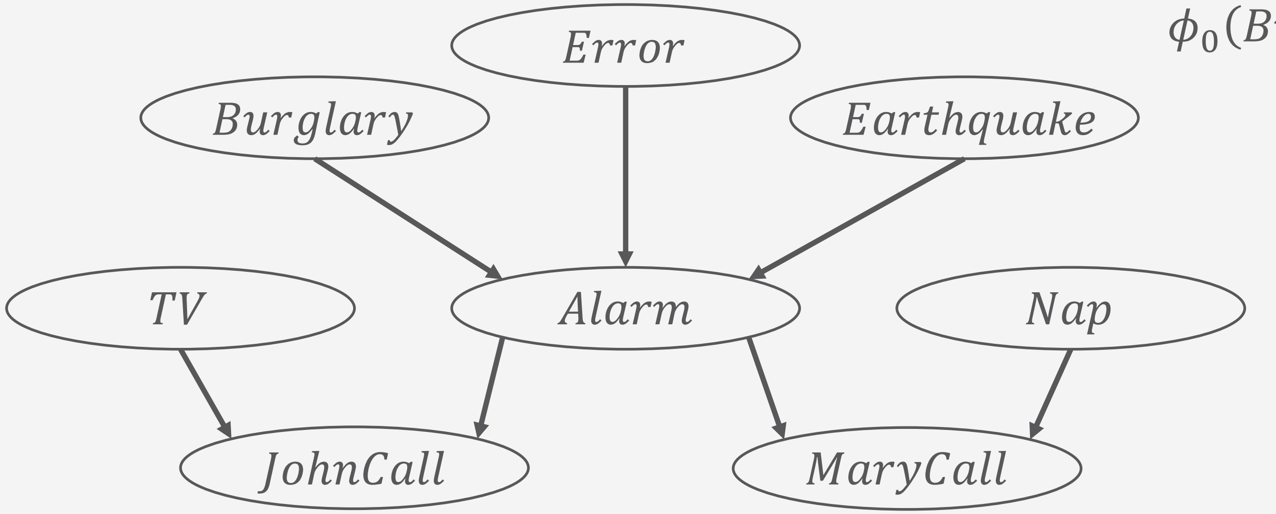


$Epid$	$Travel$	ϕ_1
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

$Epid$	$Sick$	ϕ_2
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

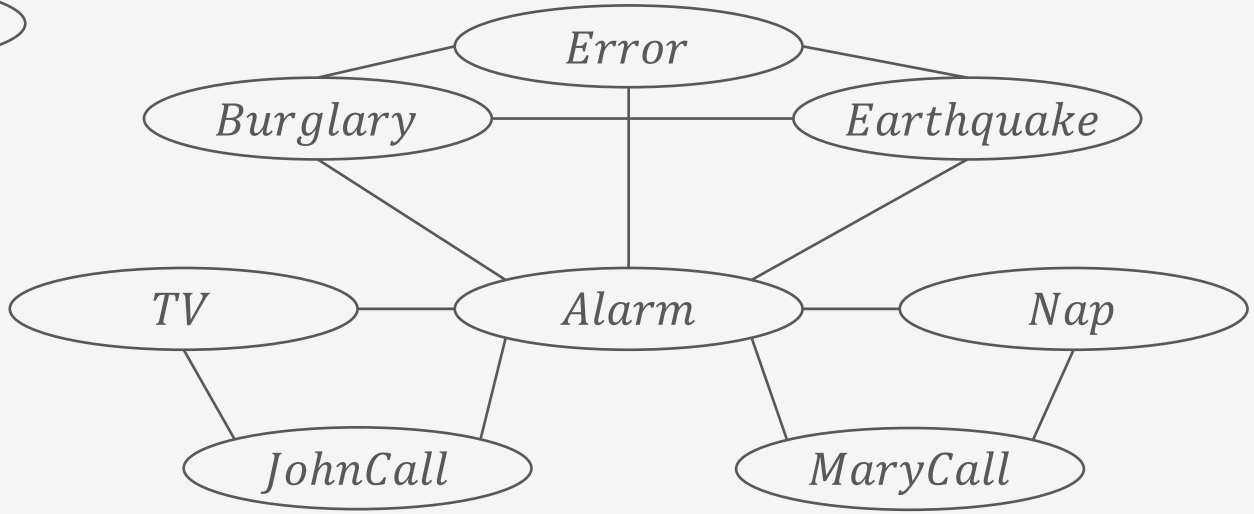
Faktormodelle: Beispiel Umwandlung BN in MN

- Ein Beispiel mit mehreren Elternknoten



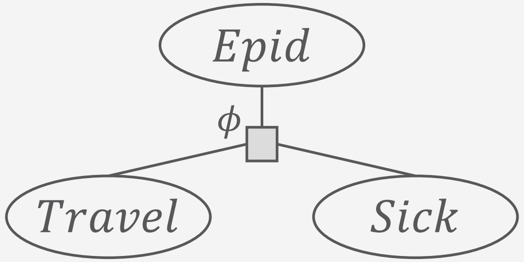
$$P(Burglary), P(Error), P(Earthquake), P(TV), P(Nap)$$
$$P(Alarm|Burglary, Error, Earthquake)$$
$$P(JohnCall \mid TV, Alarm)$$
$$P(MaryCall \mid Nap, Alarm)$$

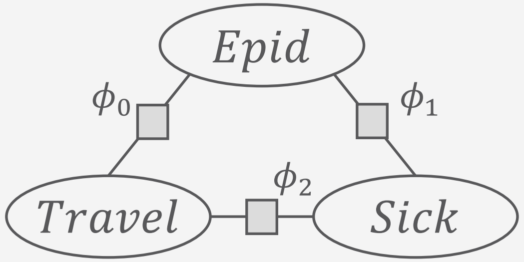
$$\phi_0(Burglary), \phi_1(Error), \phi_2(Earthquake), \phi_3(TV), \phi_4(Nap)$$
$$\phi_5(Alarm, Burglary, Error, Earthquake)$$
$$\phi_6(JohnCall, TV, Alarm)$$
$$\phi_7(MaryCall, Nap, Alarm)$$

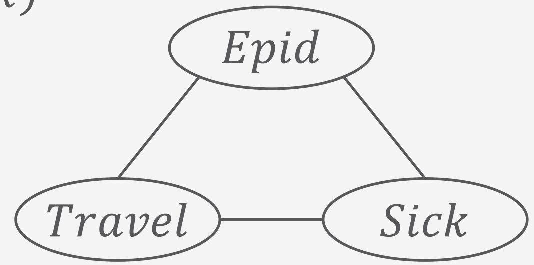


Markov Netze vs. Faktorgraphen

- Beides graphische Darstellungen eines Faktormodells
- FGs
  - Explizite graphische Darstellung der Faktoren über Faktorknoten (Boxen)
  - ❖ Erlaubt einfaches Ablesen des Modells
- MNs
  - Implizite graphische Darstellung der Faktoren über vollständige Subgraphen
    - Gegeben ein MN ohne Faktoren, Faktorisierung nicht zwangsweise eindeutig
      - Maximale Clique: $\{Epid, Travel, Sick\}$ → Faktorisierung: $P_F = \frac{1}{Z}\phi(Epid, Travel, Sick)$
      - Nicht-maximale Cliquen: $\{Epid, Travel\}, \{Epid, Sick\}, \{Travel, Sick\}$
        → Faktorisierung: $P_F = \frac{1}{Z}\phi_0(Epid, Travel) \cdot \phi_1(Epid, Sick) \cdot \phi_2(Travel, Sick)$
  - ❖ Vereinfacht die Analyse der Abhängigkeiten zwischen den Zufallsvariablen



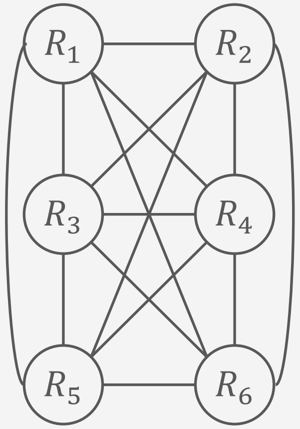




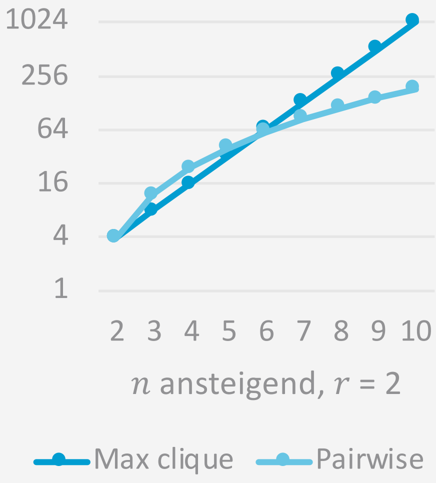
Nächstes Thema 2.D bezieht sich deshalb auf MNs
- Analoge Definitionen für FGs sind immer möglich

Minimierung der Faktorenzahl

- Gegeben ein Faktormodell $F = \{f_i\}_{i=1}^{n'}$ mit dazugehörigem MN $M = (V, E), V = \mathrm{rv}(F)$
- Anzahl n' der Faktoren in F minimieren: Nur Faktoren für maximale Cliquen zulassen
  - Wenn eine Menge von Faktoren $F' \subseteq F$ zu nicht-maximalen Cliquen einer maximalen Clique $V_C \subseteq V$ in M existiert, i.e., $\mathrm{rv}(f) \subseteq V_C$, dann Faktoren innerhalb der maximalen Clique zusammenmultiplizieren, i.e., $f_C = \prod_{f \in F'} f$
  - Nachteil: Versteckt die vorhandene Struktur
    - Kann zu einer exponentiellen Erhöhung des Speicherbedarfs führen
    - Beispiel: Vollständiger Subgraph mit n Variablen ($|\mathrm{Val}(R)| = r$ für alle Variablen) als maximale Clique, Faktoren über Zweiercliquen
      - Benötigte Einträge maximale Clique: $r^n \rightarrow$ *exponentiell* in n
      - Benötigte Einträge Zweiercliquen: $\underbrace{\frac{n(n-1)}{2}}_{\text{Anzahl Kanten}} \cdot r^2 \rightarrow$ *quadratisch* in n



Vollständiger Graph mit $n = 6$ Variablen



Zwischenzusammenfassung

- Faktormodelle
  - Faktorisierung einer vollständigen gemeinsamen Wahrscheinlichkeitsverteilung durch Faktoren
    - Faktor = positive, reell-wertige Funktion
    - Nachteil: Normalisierungskonstante nicht mehr 1
- Graphische Darstellung durch FGs oder MNs
  - FGs mit expliziten Faktoren
  - MNs mit impliziten Faktoren über Cliquen
    - Nicht eindeutig, aber einfacher zu analysieren bzgl. Unabhängigkeiten
- Umwandlung BN in Faktormodell: CPD als Faktor
  - Graphisch: von BN in FG – pro CPD ein Faktorknoten; von BN zu MN – BN moralisieren
- Beispiel: Einsatz von MNs in der Bildverarbeitung

Überblick: 2. Episodische PGMs

A. *Probabilistische Modellierung*
- Zufallsvariablen, vollständige gemeinsame Wahrscheinlichkeitsverteilung, Speicherkomplexität
- Inferenzaufgaben, Komplexität

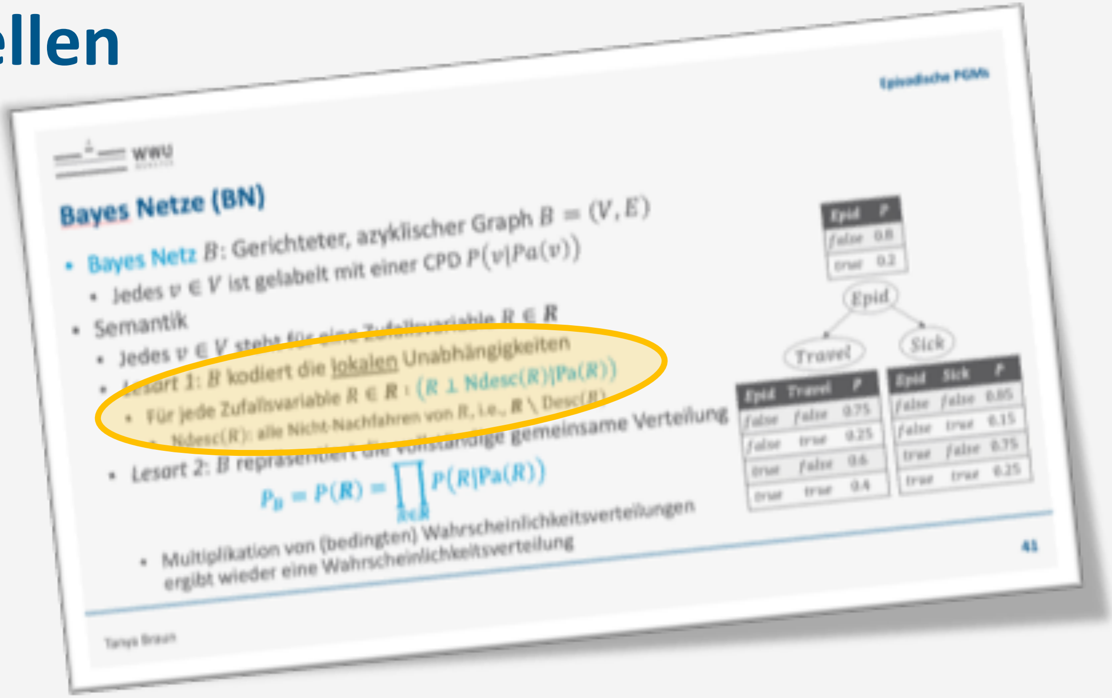
B. *Gerichtete Modelle: Bayes Netze (BNs)*
- (Bedingte) Unabhängigkeiten, Faktorisierung
- Syntax, Semantik, graphische Darstellung, Speicherkomplexität

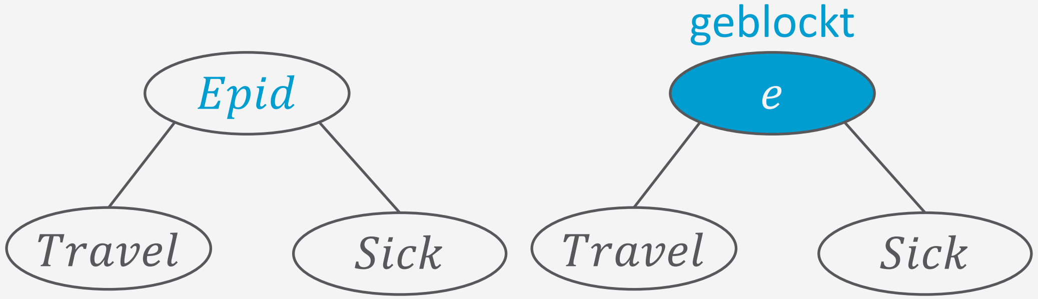
C. *Ungerichtete Modelle: Faktormodelle*
- Syntax, Semantik, graphische Darstellungen und deren Unterschiede, Speicherkomplexität
- Umwandlung von BNs zu Faktormodellen und deren graphischen Darstellungen

D. ***Unabhängigkeiten in PGMs***
- Lokale, globale und paarweise Unabhängigkeiten
- Äquivalenzbedingungen von Faktormodellen und BNs

Bedingte Unabhängigkeiten in Faktormodellen

- BNs kodieren bedingte Unabhängigkeiten
  - *Lesart 1* (<u>Lokale</u> Unabhängigkeiten): Jedes R ist unabhängig von Nicht-Nachfahren gegeben Eltern
    - Dann nur noch abhängig von Nachfahren (Kinder, Kindeskinder, …)
    - Beispiel: $Travel$ unabhängig von Nichtnachfahre $Sick$ gegeben Elternknoten $Epid$
- Wie sieht das bei Faktormodellen aus?
- Intuition
  - Einflüsse „fließen“ entlang der Pfade im MN
  - Einflüsse werden *geblockt*, wenn wir Knoten auf den Pfaden mit Beobachtungen belegen
    - Bzw. allgemeiner: auf diese Knoten konditionieren





Formale Betrachtung von bedingten Unabhängigkeiten in Faktormodellen

- Gegeben ein Faktormodell F mit MN (V, E) über Zufallsvariablen $\boldsymbol{R}$
- Lokale Unabhängigkeiten in F:
  Jedes $R \in \boldsymbol{R}$ ist unabhängig von allen anderen Zufallsvariablen gegeben seine Nachbarschaft, i.e.,
  $$\left\{\left(R \perp \boldsymbol{R}' \mid \mathrm{Nb}(R)\right) \mid R \in \boldsymbol{R} \wedge \boldsymbol{R}' = \boldsymbol{R} \setminus \{R\} \setminus \mathrm{Nb}(R)\right\}$$
  - Nachbarschaft blockiert sämtliche Pfade zu Nicht-Nachbarn
  - Nachbarschaft auch Markov Blanket (seltener: Markov Decke) genannt: $\mathrm{MB}(R) = \mathrm{Nb}(R)$
    - Analoge Definition von lokaler Unabhängigkeit: Jedes $R \in \boldsymbol{R}$ ist unabhängig von allen anderen Zufallsvariablen gegeben sein Markov Blanket, i.e.,
      $$\left\{\left(R \perp \boldsymbol{R}' \mid \mathrm{MB}(R)\right) \mid R \in \boldsymbol{R} \wedge \boldsymbol{R}' = \boldsymbol{R} \setminus \{R\} \setminus \mathrm{MB}(R)\right\}$$

- Gibt es weitere Unabhängigkeiten, die gelten?

Formale Betrachtung von bedingten Unabhängigkeiten in Faktormodellen

- Gegeben ein Faktormodell F mit MN (V, E) über Zufallsvariablen $\boldsymbol{R}$
- Paarweise Unabhängigkeiten in F:
  Jedes Paar von Nicht-Nachbarn $R_i, R_j \in \boldsymbol{R}$ ist unabhängig gegeben alle anderen Zufallsvariablen, i.e.,
  $$\left\{\left(R_i \perp R_j \mid \boldsymbol{R}'\right) \mid R_i, R_j \in \boldsymbol{R} \wedge \left\{R_i, R_j\right\} \notin E \wedge \boldsymbol{R}' = \boldsymbol{R} \setminus \left\{R_i, R_j\right\}\right\}$$
  - Wenn man auf alle verbleibenden Zufallsvariablen konditioniert, müssen sämtliche Pfade zwischen zwei Zufallsvariablen, die nicht benachbart sind, blockiert sein
  - Schwächere Aussage als die zu lokalen Unabhängigkeiten

Warum?

- Stärkere Aussage möglich?

Formale Betrachtung von bedingten Unabhängigkeiten in Faktormodellen

- Gegeben ein MN (V, E) für ein Faktormodell F über Zufallsvariablen $\boldsymbol{R}$
- Gegeben eine Menge $\boldsymbol{S} \subseteq \boldsymbol{R}$
- Aktiver Pfad
  - Ein Pfad $R_1 — \dots — R_k$ in (V, E) ist aktiv gegeben $\boldsymbol{S}$, wenn keins der $R_i, i \in \{1, \dots, k\}$ in $\boldsymbol{S}$ ist
    - Formal: $\forall\, i \in \{1, \dots, k\} : R_i \notin \boldsymbol{S}$
    - D.h., ein Pfad ist *inaktiv* gegeben $\boldsymbol{S}$, wenn mindestens ein R_i in $\boldsymbol{S}$ ist, i.e., $\exists\, i \in \{1, \dots, k\} : R_i \in \boldsymbol{S}$
- Separation (Trennung; manchmal auch *m*-Separation genannt, *m* von *M*arkov)
  - Eine Menge $\boldsymbol{S}$ separiert zwei Mengen $\boldsymbol{T}, \boldsymbol{U}$ in (V, E), wenn es keinen aktiven Pfad zwischen jeder Variable $T \in \boldsymbol{T}$ und jeder Variable $U \in \boldsymbol{U}$ gegeben $\boldsymbol{S}$ gibt, notiert als $sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S})$
    - $\boldsymbol{S}$ auch Separator genannt
    - D.h., alle Pfade zwischen den Variablen in $\boldsymbol{T}, \boldsymbol{U}$ müssen inaktiv sein gegeben $\boldsymbol{S}$
      - Auf jedem Pfad zwischen T, U muss es ein R_i geben, was in $\boldsymbol{S}$ liegt

Formale Betrachtung von bedingten Unabhängigkeiten in Faktormodellen

- Globale Unabhängigkeiten, die in (V, E) gelten:
$$\{(\boldsymbol{T} \perp \boldsymbol{U} \mid \boldsymbol{S}) \mid sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S})\}$$
- Beweis: Gegeben $sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S})$, gilt $(\boldsymbol{T} \perp \boldsymbol{U} \mid \boldsymbol{S})$?
  - Initial: $\boldsymbol{T} \cup \boldsymbol{U} \cup \boldsymbol{S} = \boldsymbol{R}$
    - Da $\boldsymbol{S}$ ein Separator zwischen $\boldsymbol{T}$ und $\boldsymbol{U}$ ist, gibt es keine direkten Kanten zwischen $\boldsymbol{T}$ und $\boldsymbol{U}$
    - Damit ist jede Clique in (V, E) (also jeder Faktor aus F) entweder in $\boldsymbol{T} \cup \boldsymbol{S}$ oder in $\boldsymbol{U} \cup \boldsymbol{S}$ enthalten
    - Mit $F_{\boldsymbol{T}}$ die Menge an Faktoren, deren Cliquen in $\boldsymbol{T} \cup \boldsymbol{S}$ liegen, und $F_{\boldsymbol{U}}$ die verbleibenden Faktoren
$$P_F = \frac{1}{Z} \prod_{f \in F} f = \frac{1}{Z} \prod_{f \in F_{\boldsymbol{T}}} f \cdot \prod_{f \in F_{\boldsymbol{U}}} f = \frac{1}{Z} f(\boldsymbol{T}, \boldsymbol{S}) \cdot g(\boldsymbol{U}, \boldsymbol{S})$$
      - Da Faktoren in $F_{\boldsymbol{T}}$ keine Zufallsvariablen aus $\boldsymbol{U}$ und Faktoren in $F_{\boldsymbol{U}}$ keine Zufallsvariablen aus $\boldsymbol{T}$ enthalten, lässt sich das Faktorprodukt aufteilen und zu jeweils einem Faktor über $\boldsymbol{T} \cup \boldsymbol{S}$ und $\boldsymbol{U} \cup \boldsymbol{S}$ zusammenfassen
    - Mit der Faktorisierung $f(\boldsymbol{T}, \boldsymbol{S}) \cdot g(\boldsymbol{U}, \boldsymbol{S})$ gilt $(\boldsymbol{T} \perp \boldsymbol{U} \mid \boldsymbol{S})$ (Bayes Theorem, Definition bedingte Unabhängigkeit)

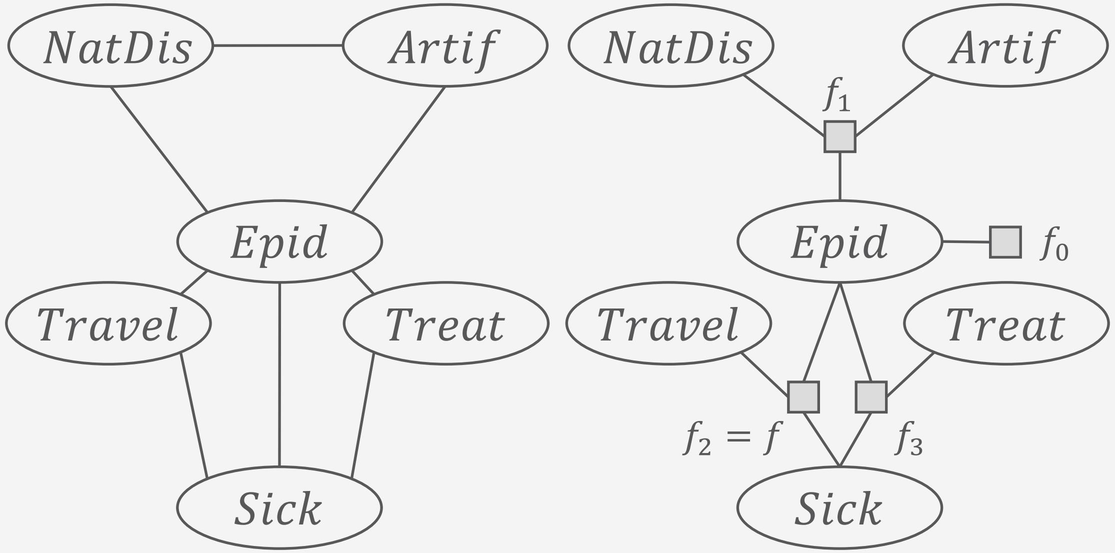
Formale Betrachtung von bedingten Unabhängigkeiten in Faktormodellen

- Globale Unabhängigkeiten, die in (V, E) gelten:
$$\{(\boldsymbol{T} \perp \boldsymbol{U} \mid \boldsymbol{S}) \mid sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S})\}$$
- Beweis: Gegeben $sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S})$, gilt $(\boldsymbol{T} \perp \boldsymbol{U} \mid \boldsymbol{S})$?
  - Allgemein: $\boldsymbol{T} \cup \boldsymbol{U} \cup \boldsymbol{S} \subset \boldsymbol{R}$ mit $\boldsymbol{V} = \boldsymbol{R} \setminus (\boldsymbol{T} \cup \boldsymbol{U} \cup \boldsymbol{S})$
    - $\boldsymbol{V}$ lässt sich in zwei Mengen $\boldsymbol{V}_1, \boldsymbol{V}_2$ aufteilen, so dass $\boldsymbol{S}$ Separator zwischen $\boldsymbol{T} \cup \boldsymbol{V}_1$ und $\boldsymbol{U} \cup \boldsymbol{V}_2$ ist
    - Danach gilt die obige Argumentation:
      Mit $F_{\boldsymbol{T}}$ die Menge an Faktoren, deren Cliquen in $\boldsymbol{T} \cup \boldsymbol{V}_1 \cup \boldsymbol{S}$ liegen, und $F_{\boldsymbol{U}}$ die verbleibenden Faktoren
$$P_F = \frac{1}{Z} \prod_{f \in F} f = \frac{1}{Z} \prod_{f \in F_{\boldsymbol{T}}} f \cdot \prod_{f \in F_{\boldsymbol{U}}} f = \frac{1}{Z} f(\boldsymbol{T}, \boldsymbol{V}_1, \boldsymbol{S}) \cdot g(\boldsymbol{U}, \boldsymbol{V}_2, \boldsymbol{S})$$
    - Mit Dekompositionsregel bedingter Unabhängigkeiten, $(\boldsymbol{A} \perp \boldsymbol{B}, \boldsymbol{C} \mid \boldsymbol{D}) \Rightarrow (\boldsymbol{A} \perp \boldsymbol{B} \mid \boldsymbol{D})$, gilt dann $(\boldsymbol{T} \perp \boldsymbol{U} \mid \boldsymbol{S})$ mit dem gleichen Argument wie bei der initialen Betrachtung
- *Monoton*: falls $sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S})$ gilt, dann gilt auch $sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S}')$ für $\boldsymbol{S}' \supseteq \boldsymbol{S}$

Beispiel: Unabhängigkeiten in Faktormodellen

- Paarweise Unabhängigkeiten:
  - $NatDis \perp Travel \mid \{Artif, Epid, Treat, Sick\}$
    - $NatDis$ mit $Artif$ austauschbar
    - $Travel$ mit $Treat, Sick$ austauschbar
  - $Treat \perp Travel \mid \{NatDis, Artif, Epid, Sick\}$
  - $Epid$ hat keine Nichtnachbarn
- Lokale Unabhängigkeiten
  - $NatDis \perp \{Travel, Treat, Sick\} \mid \{Artif, Epid\}$
    - $NatDis$ mit $Artif$ austauschbar
  - $Treat \perp \{NatDis, Artif, Travel\} \mid \{Epid, Sick\}$
    - $Treat$ mit $Travel$ austauschbar
  - $Sick \perp \{NatDis, Artif\} \mid \{Epid, Travel, Treat\}$

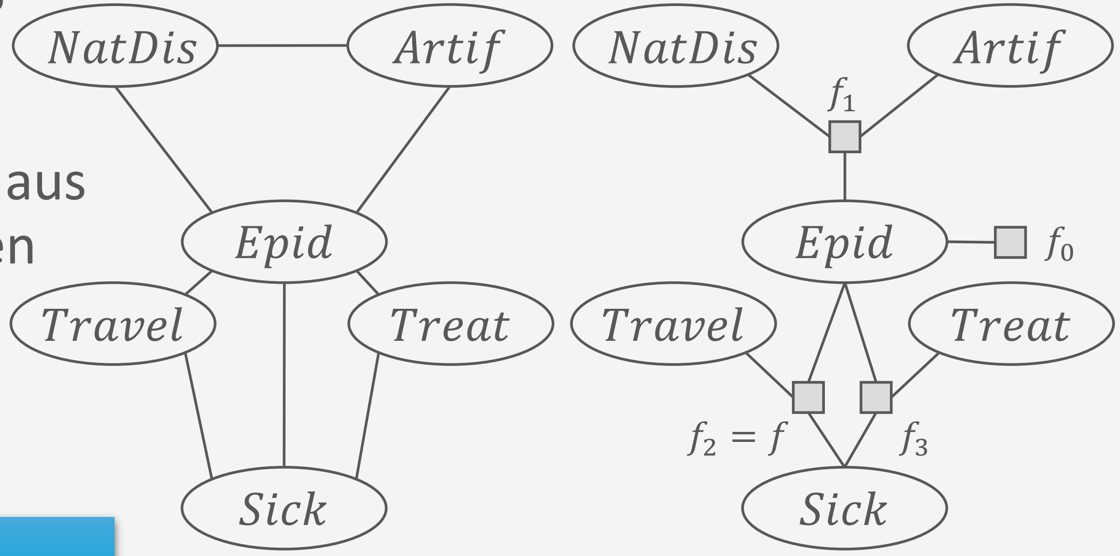
Paarweise: $\left(R_i \perp R_j \mid \boldsymbol{R}'\right), \boldsymbol{R}' = \boldsymbol{R} \setminus \{R_i, R_j\}$
Lokal: $\left(R \perp \boldsymbol{R}' \mid \mathrm{MB}(R)\right), \boldsymbol{R}' = \boldsymbol{R} \setminus \mathrm{MB}(R) \setminus \{R\}$
Global: $\{(\boldsymbol{T} \perp \boldsymbol{U} \mid \boldsymbol{S}) \mid sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S})\}$



Beispiel: Unabhängigkeiten in Faktormodellen

- Globale Unabhängigkeiten:
  - Paarweise und lokale Unabhängigkeiten lassen sich auf $\{(\boldsymbol{T} \perp \boldsymbol{U} \mid \boldsymbol{S}) \mid sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S})\}$ abbilden
  - $\{NatDis, Artif\} \perp \{Travel, Treat, Sick\} \mid Epid$
    - Dekomposition: Untermengen von $\{NatDis, Artif\}$ bzw. $\{Travel, Treat, Sick\}$ möglich
      - Monoton → Menge rechts vom | erweiterbar
  - Menge der „repräsentativen“ Unabhängigkeiten, aus denen wir alle anderen Unabhängigkeiten ableiten können:
    - $\{NatDis, Artif\} \perp \{Travel, Treat, Sick\} \mid Epid$
    - $Treat \perp Travel \mid \{Epid, Sick\}$

Paarweise: $(R_i \perp R_j \mid \boldsymbol{R}'), \boldsymbol{R}' = \boldsymbol{R} \setminus \{R_i, R_j\}$
Lokal: $(R \perp \boldsymbol{R}' \mid \mathrm{MB}(R)), \boldsymbol{R}' = \boldsymbol{R} \setminus \mathrm{MB}(R) \setminus \{R\}$
Global: $\{(\boldsymbol{T} \perp \boldsymbol{U} \mid \boldsymbol{S}) \mid sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S})\}$



Unabhängigkeit gegeben $Epid$ vereinfacht die paarweise Unabhängigkeit
$Treat \perp Travel \mid \{NatDis, Artif, Epid, Sick\}$

Fortsetzung: Unabhängigkeiten in BNs

- BNs kodieren lokale Unabhängigkeiten
  - Für jede Zufallsvariable $R \in \boldsymbol{R} : \big(R \perp \mathrm{Ndesc}(R) \mid \mathrm{Pa}(R)\big)$

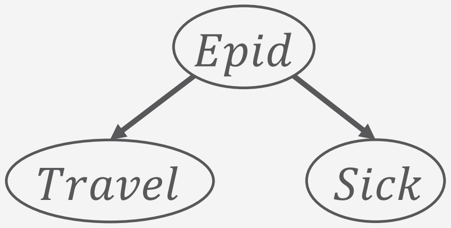
vs.

- MNs kodieren lokale Unabhängigkeiten
  - Für jede Zufallsvariable $R \in \boldsymbol{R} : \big(R \perp \boldsymbol{R}' \mid \mathrm{Nb}(R)\big)$

Was ist der Unterschied?

- Woher kommt der Unterschied?
  - Einflüsse laufen auch entlang der Kanten, aber was ist anders?
- Welche weiteren bzw. analogen Unabhängigkeiten gelten in BNs?

$Epid$	P
$false$	0.8
$true$	0.2

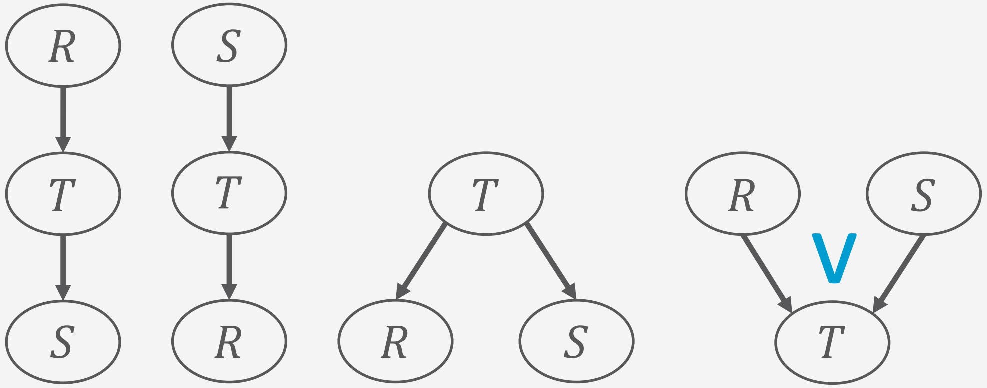


$Epid$	$Travel$	P
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

$Epid$	$Sick$	P
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

Einfluss von Zufallsvariablen in BNs

- Direkte Verbindung: direkter Einfluss
- Indirekte Verbindung: Fallunterscheidung
  a. Indirekter kausaler Effekt
  b. Indirekter Evidenzeffekt
  c. Gemeinsame Ursache
  d. Gemeinsamer Effekt → v-Struktur



a. b. c. d.

- Gegeben durch lokale Unabhängigkeiten
  - Fall a-c: R beinflusst S via T, aber nur wenn T nicht beobachtet
  - Fall d: Wenn T nicht beobachtet, sind R und S unabhängig
    - Wenn T beobachtet, sind R und S abhängig
      - Wert z muss durch R oder S zustande kommen
  - Aussagen gelten auch in größeren Netzen mit mehr Knoten auf Pfad zwischen R, S
    - Fall d: weder T noch Nachfahre beobachtet

Aussagen über BNs analog zu Faktormodellen bezogen auf paarweise, lokale und globale Unabhängigkeiten müssen deshalb *v-Strukturen* miteinbeziehen.

Zurück zu den bisherigen Unabhängigkeiten in BNs

- BNs kodieren lokale Unabhängigkeiten
  - Für jede Zufallsvariable $R \in \boldsymbol{R} : \left(R \perp \mathrm{Ndesc}(R) \mid \mathrm{Pa}(R)\right)$

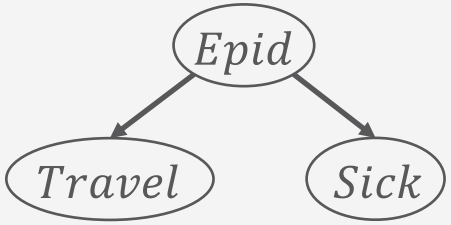
vs.

- MNs kodieren lokale Unabhängigkeiten
  - Für jede Zufallsvariable $R \in \boldsymbol{R} : \left(R \perp \boldsymbol{R}' \mid \mathrm{Nb}(R)\right)$

Was ist der Unterschied?

- Woher kommt der Unterschied?
  - Wenn Eltern und Kinder als Nachbarn von R gegeben wären, werden weitere Elternknoten der Kindsknoten abhängig von R
- Welche weiteren bzw. analogen Unabhängigkeiten gelten in BNs?

$Epid$	P
$false$	0.8
$true$	0.2



$Epid$	$Travel$	P
$false$	$false$	0.75
$false$	$true$	0.25
$true$	$false$	0.6
$true$	$true$	0.4

$Epid$	$Sick$	P
$false$	$false$	0.85
$false$	$true$	0.15
$true$	$false$	0.75
$true$	$true$	0.25

Formale Betrachtung von bedingten Unabhängigkeiten in BNs

- Gegeben ein BN B über Zufallsvariablen $\boldsymbol{R}$
- Markov Blanket in BNs:
  Eltern, Kinder und alle weiteren Eltern der Kinder, i.e.,

$$\mathrm{MB}(R) = \mathrm{Pa}(R) \cup \mathrm{Ch}(R) \cup \boxed{\bigcup_{R' \in \mathrm{Ch}(R)} \mathrm{Pa}(R') \setminus \{R\}}$$

Fängt alle *v-Strukturen* ab

- Lokale Unabhängigkeiten in B *analog zu einem MN*:
  Jedes $R \in \boldsymbol{R}$ ist unabhängig von allen anderen Zufallsvariablen gegeben sein Markov Blanket, i.e.,

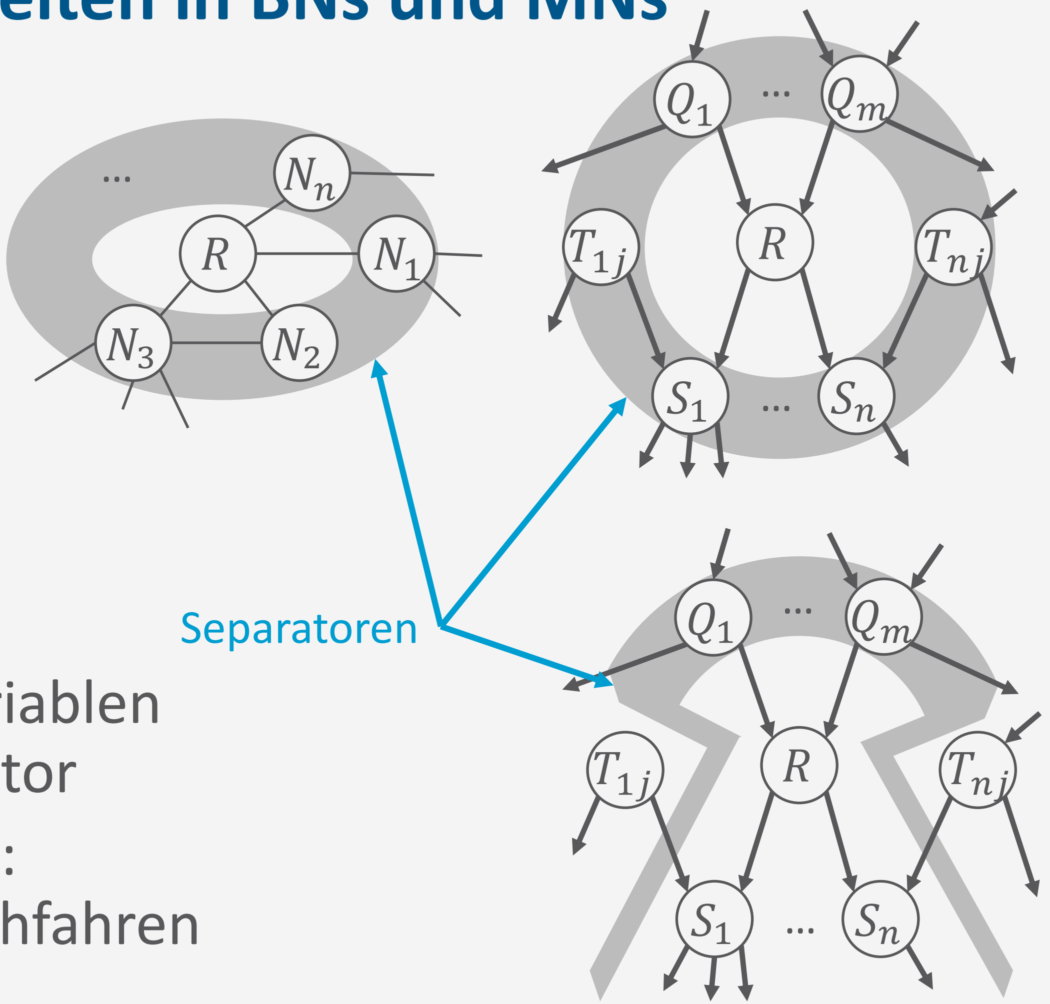
$$\left\{ \left(R \perp \boldsymbol{R}' \mid \mathrm{MB}(R) \right) \mid R \in \boldsymbol{R} \wedge \boldsymbol{R}' = \boldsymbol{R} \setminus \{R\} \setminus \mathrm{MB}(R) \right\}$$

  - Markov Blanket blockiert sämtliche Pfade, inklusive v-Strukturen

Veranschaulichung der lokalen Unabhängigkeiten in BNs und MNs

- Markov Blanket einer Zufallsvariable R
  - in einem Faktormodell:
    - Menge der direkten Nachbarn N_i von R im MN
  - in einem BN:
    - Vereinigung der Mengen
      - Eltern Q_k von R
      - Kinder S_i von R

      Nachbarn
      - Eltern T_{ij} der Kinder S_i, die nicht R sind
- R bedingt unabhängig von allen anderen Zufallsvariablen im Modell gegeben sein Markov Blanket als Separator
- Vorherige Definition lokaler Unabhängigkeit in BNs: Gegeben die Eltern als Separatoren: Alle Nicht-Nachfahren



Formale Betrachtung von bedingten Unabhängigkeiten in BNs

- Gegeben ein BN B über Zufallsvariablen $\boldsymbol{R}$
- Paarweise Unabhängigkeiten in B:
  - In der Literatur werden für BNs normalerweise keine paarweise Unabhängigkeiten betrachtet
    - Paarweise Unabhängigkeiten in MN: Jedes Paar von Nicht-Nachbarn $R_i, R_j \in \boldsymbol{R}$ ist unabhängig gegeben alle anderen Zufallsvariablen, i.e.,
$$\left\{\left(R_i \perp R_j \mid \boldsymbol{R}'\right) \mid R_i, R_j \in \boldsymbol{R} \wedge \left\{R_i, R_j\right\} \notin E \wedge \boldsymbol{R}' = \boldsymbol{R} \setminus \left\{R_i, R_j\right\}\right\}$$
    - Kann so nicht definiert werden, da bei gemeinsamen Nachfahren R_i, R_j abhängig werden, da der Nachfahre in $\boldsymbol{R}'$ zwangsläufig vorkommt
  - Man könnte eine Aussage zu Nicht-Nachbarn treffen, wenn man
    - nur Zufallsvariablen R_i, R_j betrachtet, die keine gemeinsamen Nachfahren haben
    - die gemeinsamen Nachfahren aus $\boldsymbol{R}'$ ausschließt

Formale Betrachtung von bedingten Unabhängigkeiten in BNs

Fängt Abhängigkeiten durch *v-Strukturen* ein

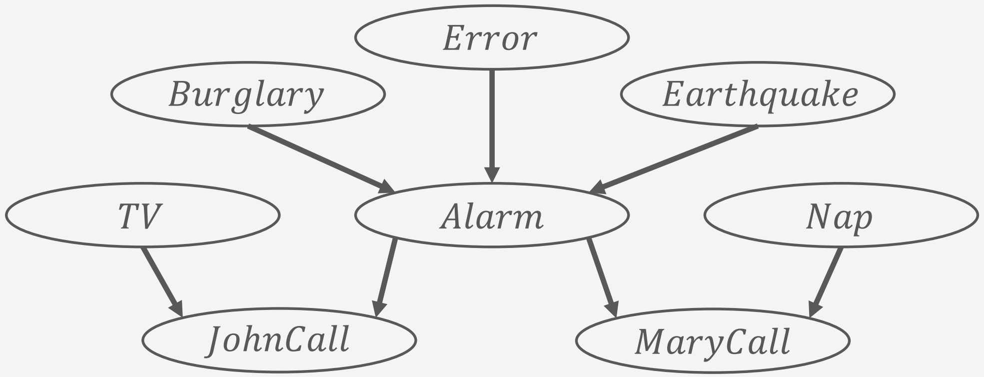
- Gegeben disjunkte Mengen $\boldsymbol{S}, \boldsymbol{T}, \boldsymbol{U}$ von Zufallsvariablen in B
- Aktive Verbindung zwischen zwei Zufallsvariablen T, U
  - Verbindung: *ungerichteter* Pfad zwischen T, U
  - Verbindung zwischen zwei Zufallsvariablen T, U ist aktiv gegeben $\boldsymbol{S}$, wenn
    - Falls wir eine v-Struktur $R_{i-1} \rightarrow R_i \leftarrow R_{i+1}$ auf der Verbindung haben, dann ist R_i oder ein Nachfahre von R_i in $\boldsymbol{S}$
    - Kein weiterer Knoten auf der Verbindung ist in $\boldsymbol{S}$
- Gerichtete Separation (*directed separation, d-separation*)
  - $\boldsymbol{T}$ und $\boldsymbol{U}$ sind d-separiert gegeben $\boldsymbol{S}$, wenn es keine aktive Verbindung zwischen jeder Variable $T \in \boldsymbol{T}$ und jeder Variable $U \in \boldsymbol{U}$ gegeben $\boldsymbol{S}$ gibt, notiert als $d\text{-}sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S})$
- Globale Unabhängigkeiten, die in B bzw. P_B gelten*

$$\{(\boldsymbol{T} \perp \boldsymbol{U} \mid \boldsymbol{S}) \mid d\text{-}sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S})\}$$

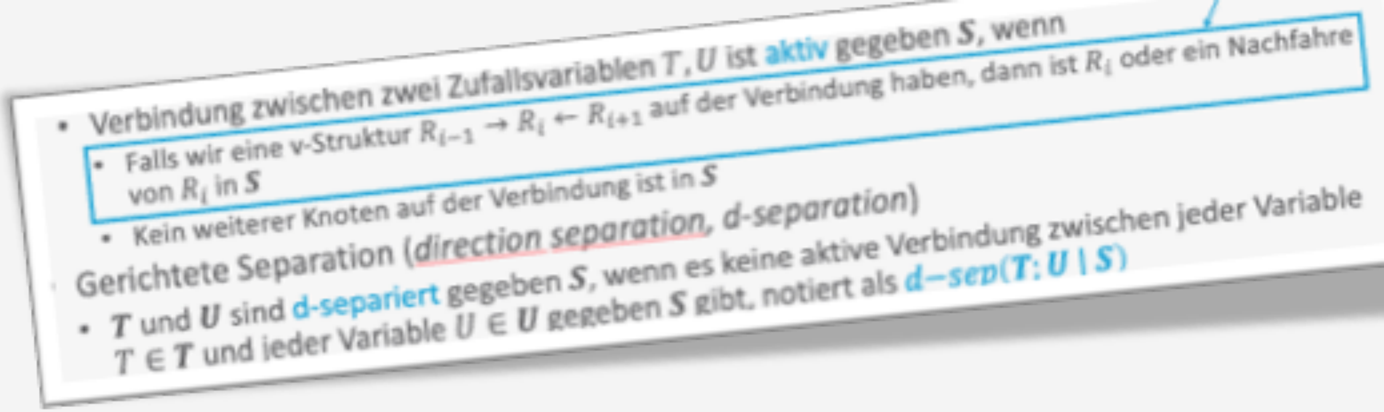
*Ohne Beweis
Für Interessierte: Abschnitt 4.5.1 im PGM Buch

Beispiele: Unabhängigkeiten in BNs

- Lokale Unabhängigkeiten
  - $Burglary \perp \{TV, JohnCall, Nap, MaryCall\} \mid \overbrace{\{Alarm, Error, Earthquake\}}^{\mathrm{MB}(Burglary)}$
    - $Burglary$ austauschbar mit $Error, Earthquake$
  - $TV \perp \{Burglary, Error, Earthquake, Nap, MaryCall\} \mid \{JohnCall, Alarm\}$
    - TV zusammen mit $JohnCall$ austauschbar mit Nap zusammen mit $MaryCall$
  - $JohnCall \perp \{Burglary, Error, Earthquake, Nap, MaryCall\} \mid \{TV, Alarm\}$
    - $JohnCall$ zusammen mit TV austauschbar mit $MaryCall$ zusammen mit Nap
  - Zu $Alarm$ gibt es keine lokalen Unabhängigkeiten
    - $\mathrm{MB}(Alarm)$ beinhaltet alle anderen Knoten
      - Eltern: $Burglary, Error, Earthquake$
      - Kinder: $JohnCall, MaryCall$
      - Weitere Eltern von Kindern: TV, Nap



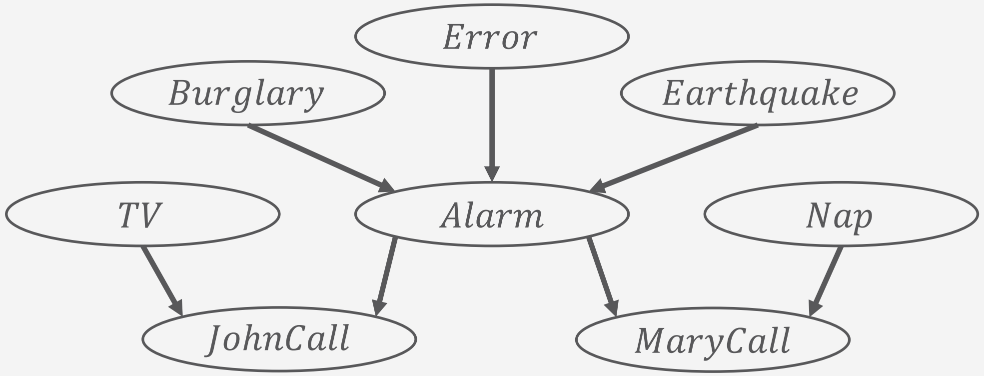
Beispiele: Unabhängigkeiten in BNs


- Verbindung zwischen zwei Zufallsvariablen T, U ist aktiv gegeben $\boldsymbol{S}$, wenn
  - Falls wir eine v-Struktur $R_{i-1} \rightarrow R_i \leftarrow R_{i+1}$ auf der Verbindung haben, dann ist R_i oder ein Nachfahre von R_i in $\boldsymbol{S}$
  - Kein weiterer Knoten auf der Verbindung ist in $\boldsymbol{S}$

Gerichtete Separation (direction separation, d-separation)
- $\boldsymbol{T}$ und $\boldsymbol{U}$ sind d-separiert gegeben $\boldsymbol{S}$, wenn es keine aktive Verbindung zwischen jeder Variable $T \in \boldsymbol{T}$ und jeder Variable $U \in \boldsymbol{U}$ gegeben $\boldsymbol{S}$ gibt, notiert als $d\text{-}sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S})$

- Globale Unabhängigkeiten
  - Lokale Unabhängigkeiten lassen sich auf $\{(\boldsymbol{T} \perp \boldsymbol{U} \mid \boldsymbol{S}) \mid d{-}sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S})\}$ abbilden
    - Beispiel: $Burglary \perp \{TV, JohnCall, Nap, MaryCall\} \mid \{Alarm, Error, Earthquake\}$
      - $\boldsymbol{T} = \{Burglary\}, \boldsymbol{U} = \{TV, JohnCall, Nap, MaryCall\}, \boldsymbol{S} = \{Alarm, Error, Earthquake\}$
      - Pfad von $Burglary$ nach $JohnCall$ bzw. $MaryCall$ inaktiv, da durch $Alarm$ geblockt
      - Pfad von $Burglary$ nach TV bzw. Nap inaktiv, da durch $Alarm$ geblockt
  - $\{Burglary, Error, Earthquake\} \perp \{TV, JohnCall, Nap, MaryCall\} \mid Alarm$
    - Untermengen von $\{Burglary, Error, Earthquake\}$ bzw. $\{TV, JohnCall, Nap, MaryCall\}$ möglich
      - Nur nicht einer der anderen Mengen hinzufügen
  - Liste nicht erschöpfend

Wann wäre der Pfad aktiv?



Formale Betrachtung von Unabhängigkeiten

- Verhältnis zwischen den Unabhängigkeiten:
  - Paarweise: $\left(R_i \perp R_j \mid \boldsymbol{R}'\right), \boldsymbol{R}' = \boldsymbol{R} \setminus \left\{R_i, R_j\right\}$ in MNs
  - Lokal: $\left(R \perp \boldsymbol{R}' \mid \mathrm{MB}(R)\right), \boldsymbol{R}' = \boldsymbol{R} \setminus \mathrm{MB}(R) \setminus \{R\}$
  - Global: $\{(\boldsymbol{T} \perp \boldsymbol{U} \mid \boldsymbol{S}) \mid sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S})\}$ bzw. $\{(\boldsymbol{T} \perp \boldsymbol{U} \mid \boldsymbol{S}) \mid d\text{-}sep(\boldsymbol{T}; \boldsymbol{U} \mid \boldsymbol{S})\}$
  - ❖ Äquivalent in strikt positiven Verteilungen
    - Strikt positiv: $P(\omega) > 0$ für alle $\omega \in \mathrm{Val}(\boldsymbol{R})$
    - Andernfalls könnte eine bedingte Wahrscheinlichkeit nicht definiert sein, weil der Teiler 0 ist
- Paarweise, lokale und globale Unabhängigkeiten werden auch als paarweise, lokale und globale Markov Eigentschaft bezeichnet, die eine Verteilung erfüllt (oder eben nicht)

Formale Betrachtung von Unabhängigkeiten

- Hammersley-Clifford Theorem
  - P erfüllt eine der Markov Eigenschaften (und damit alle) in Bezug auf einen ungerichteten Graphen *genau dann, wenn* P über die Cliquen des Graphen faktorisiert werden kann
    - Wenn P eine Markov Eigenschaft erfüllt, dann lässt sich P über die Cliquen faktorisieren
    - Wenn wir die Faktorisierung von P über Cliquen darstellen, dann erfüllt der Graph die Markov Eigenschaften
- Das Theorem besagt nicht, dass die Faktorisierung eindeutig ist, und gibt auch keine Hinweise darauf, wie man eine Faktorisierung findet

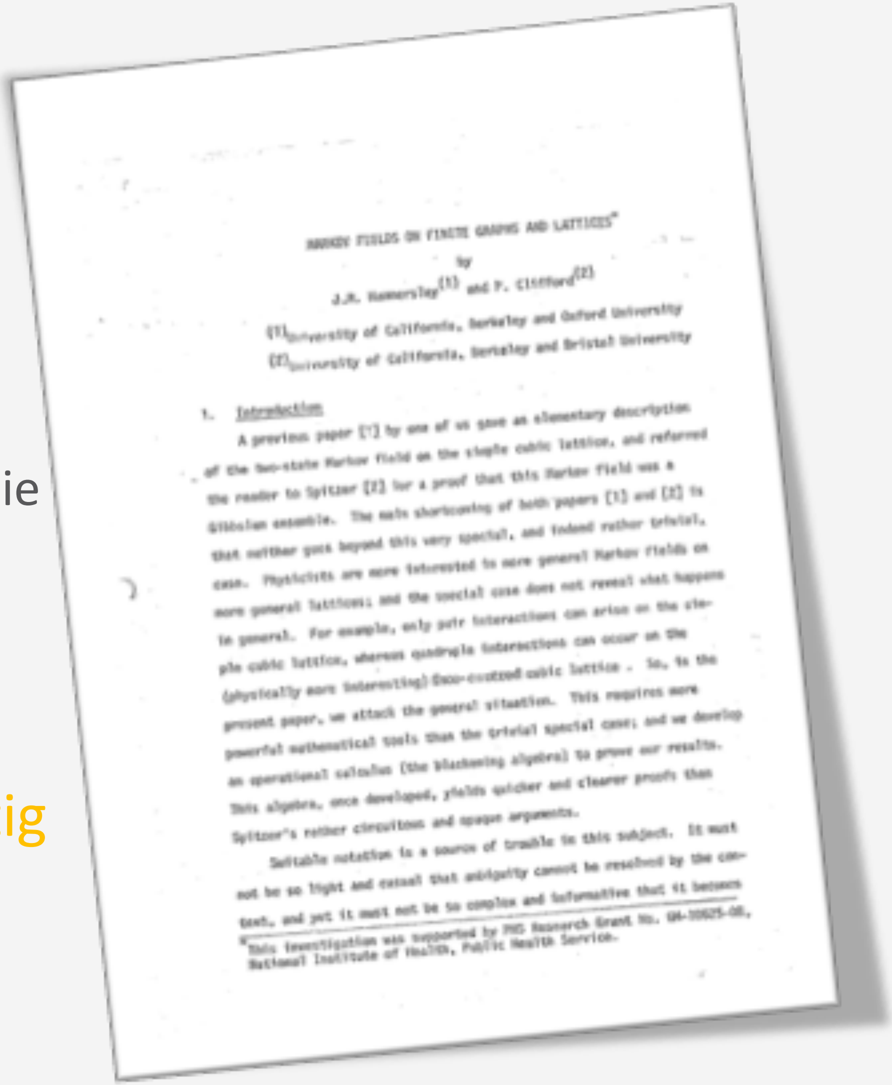


Abbildung: Screenshot der ersten Seite des Manuskripts von Hammersley und Clifford
Referenz: John Hammersley and Peter Clifford: Markov Fields on Finite Graphs and Lattices. *Unpublished*, 1971.

Von Verteilung zu Faktorisierung über Unabhängigkeiten

- Gegeben eine vollständige gemeinsame Verteilung $P_{\boldsymbol{R}}$ über Zufallsvariablen $\boldsymbol{R}$
- Gegeben einen maximalen Grad d
- Mit vollständigem Graphen starten
- Nach und nach Kanten löschen:
  - Finde ein $\boldsymbol{U}$, so dass gilt $R_i \perp R_j \mid \boldsymbol{U}$
    - $|\boldsymbol{U}|$ maximal d
      - Suche beschränken
  - Dann lösche Kante zwischen R_i, R_j

- Gegeben eine vollständige gemeinsame Verteilung $P_{\boldsymbol{R}}$ über Zufallsvariablen $\boldsymbol{R}$
- Mit leerem Graphen starten
- Minimales Markov Blanket für jedes R finden, Kanten zwischen R und $\mathrm{MB}(R)$ hinzufügen

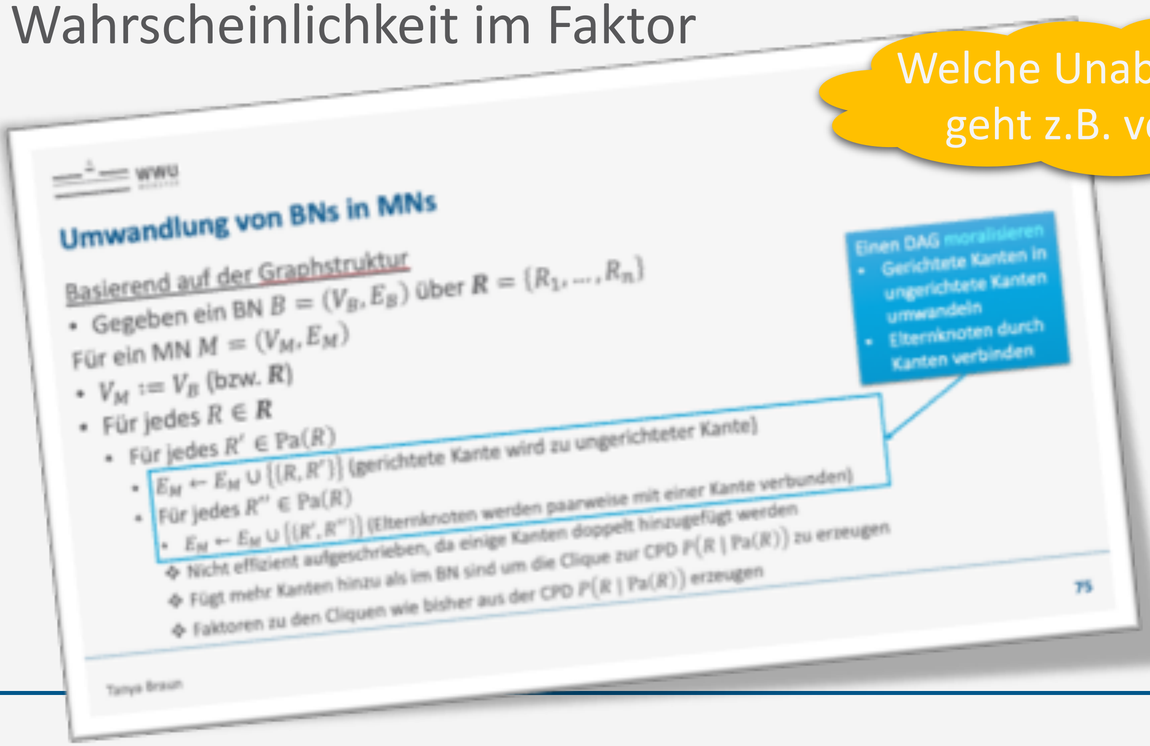
In der Praxis meistens erstmal nur Daten gegeben
- Statt über $P_{\boldsymbol{R}}$ zu gehen, versucht man gleich eine Faktorisierung zu lernen, um die kombinatorische Explosion in $P_{\boldsymbol{R}}$ zu vermeiden

→ Thema 5: Lernalgorithmen in episodischen PGMs

Für Interessierte:
Nähere Infos im PGM-Buch, Abschnitt 4.3.3 für MNs und Abschnitt 3.4 für BNs

Wiederbetrachtung: Umwandlung von BNs

- Konsequenzen der Umwandlung
  - In der graphischen Darstellung: keine Kantenrichtungen
  - In der CPDs/Faktoren: Verstecken der bedingten Wahrscheinlichkeit im Faktor



Umwandlung von BNs in MNs

Basierend auf der Graphstruktur
- Gegeben ein BN $B = (V_B, E_B)$ über $\boldsymbol{R} = \{R_1, \ldots, R_n\}$

Für ein MN $M = (V_M, E_M)$
- $V_M := V_B$ (bzw. $\boldsymbol{R}$)
- Für jedes $R \in \boldsymbol{R}$
  - Für jedes $R' \in \text{Pa}(R)$
    - $E_M \leftarrow E_M \cup \{(R, R')\}$ (gerichtete Kante wird zu ungerichteter Kante)
    - Für jedes $R'' \in \text{Pa}(R)$
      - $E_M \leftarrow E_M \cup \{(R', R'')\}$ (Elternknoten werden paarweise mit einer Kante verbunden)

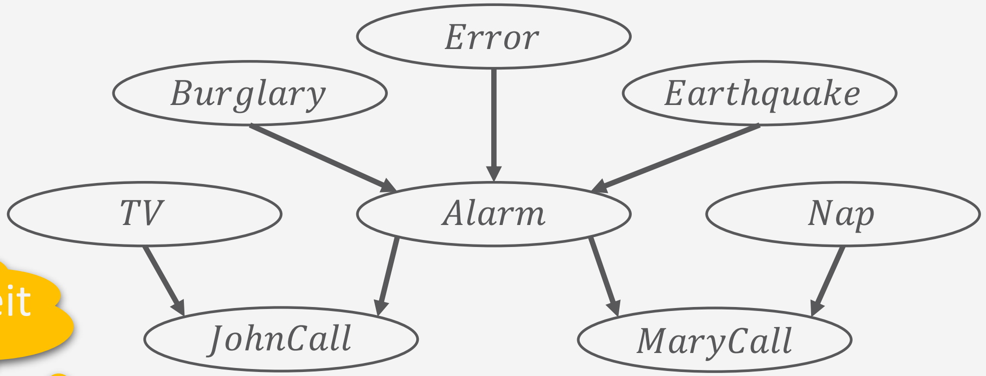
Einen DAG moralisieren
- Gerichtete Kanten in ungerichtete Kanten umwandeln
- Elternknoten durch Kanten verbinden

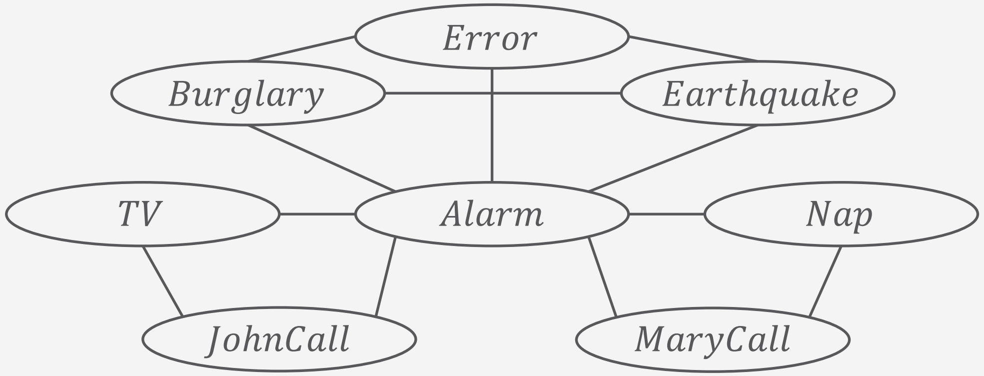
- Nicht effizient aufgeschrieben, da einige Kanten doppelt hinzugefügt werden
- Fügt mehr Kanten hinzu als im BN sind um die Clique zur CPD $P(R \mid \text{Pa}(R))$ zu erzeugen
- Faktoren zu den Cliquen wie bisher aus der CPD $P(R \mid \text{Pa}(R))$ erzeugen

75

Tanya Braun

Welche Unabhängigkeit geht z.B. verloren?



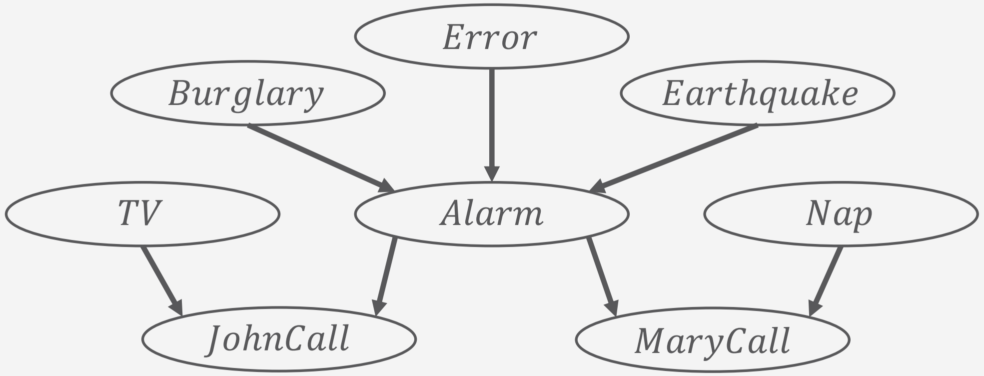


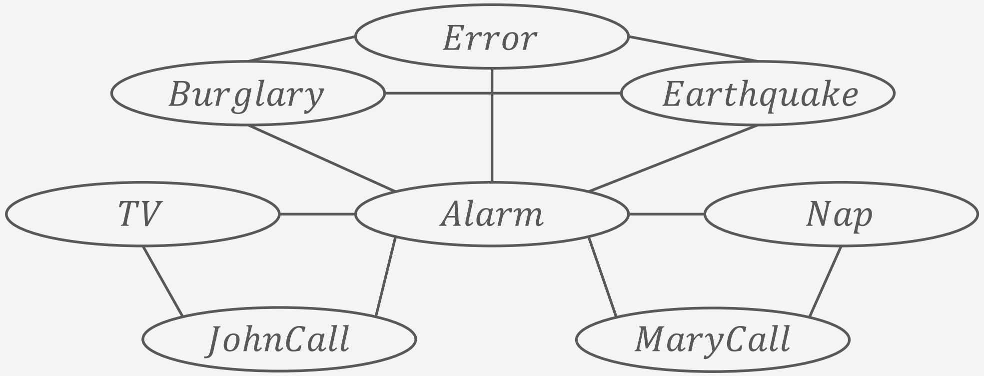
Wiederbetrachtung: Umwandlung von BNs

- Globale Unabhängigkeiten im BN, aber nicht im MN
  - Unabhängigkeit von $Burglary, Error, Earthquake$, wenn $Alarm, JohnCall, MaryCall$ nicht gegeben sind
  - Unabhängigkeit von $Alarm, TV$, wenn $JohnCall$ nicht gegeben ist, bzw. von $Alarm, Nap$, wenn $MaryCall$ nicht gegeben ist

→ Unabhängigkeiten im Rahmen von v-Strukturen gehen beim Übertragen von BNs auf Faktormodelle verloren

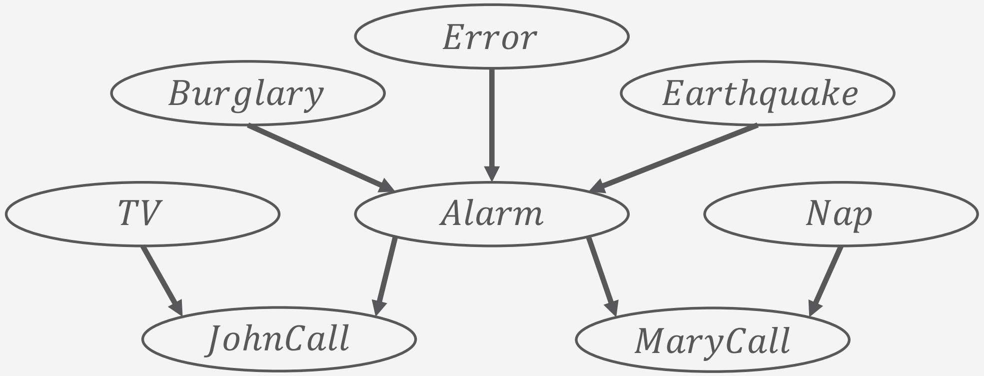
Bei allen v-Strukturen?

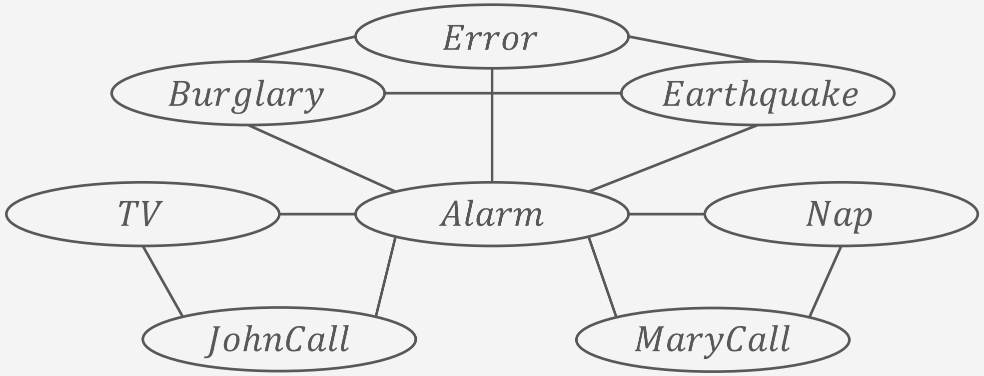




Wiederbetrachtung: Umwandlung von BNs

- Unabhängigkeiten gehen verloren, wenn beim Moralisieren eines BNs Kanten hinzugefügt werden
- BN $B = (V, E)$ ist ein moraler Graph, wenn alle Elternpaare eines Kindes durch eine Kante verbunden sind
  - Formal: für jedes $R_i, R_j \in V$ mit $R_i, R_j \in \mathrm{Pa}(R_k)$ für ein $R_k \in V$ gilt: $\left(R_i, R_j\right) \in E \vee \left(R_j, R_i\right) \in E$
  - Dann werden beim Moralisieren keine Kanten hinzugefügt
- Wenn ein BN $B = (V, E)$ ein *moraler* Graph ist, dann existiert ein Faktormodell F mit einem MN (V, E'), so dass die Unabhängigkeiten gleich sind

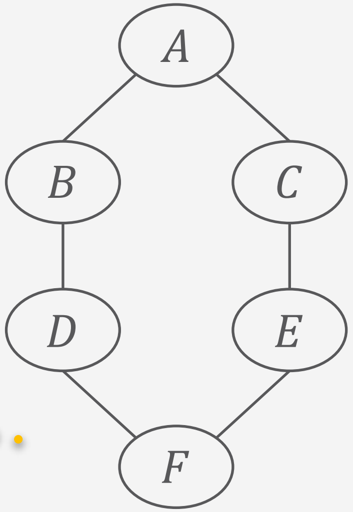


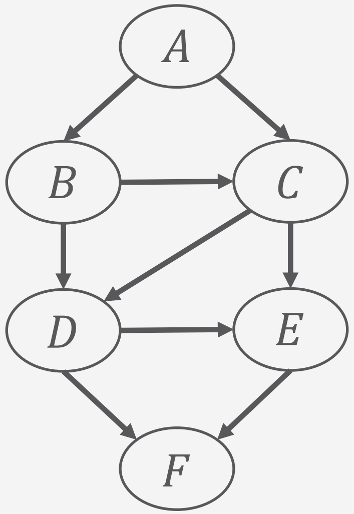


Die andere Richtung: Umwandlung von MNs in BNs

- Informell: Gegeben einer Knotenreihenfolge nach und nach Knoten in ein BN einfügen und Kanten hinzufügen, wann immer im MN eine Abhängigkeit zwischen den schon einfügten Knoten herrscht
  - Beispiel: A, B, C, D, E, F
    - A hinzufügen
    - B hinzufügen, B abhängig von A in MN → Kante
    - C hinzufügen, C abhängig von A in MN → Kante
      - Aber C auch abhängig von B durch E, F, D → Kante
    - D abhängig von B, aber durch F, E auch von C
    - E abhängig von C, aber durch F auch von D
    - F abhängig von D, E

Wann fügen wir keine Kanten ein?

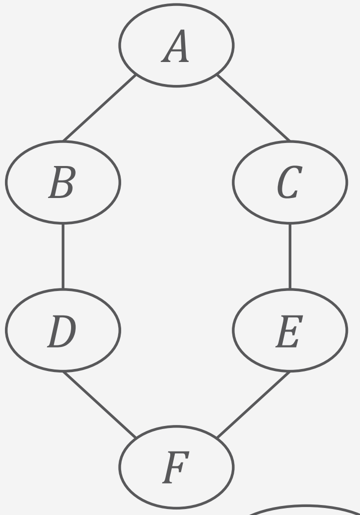


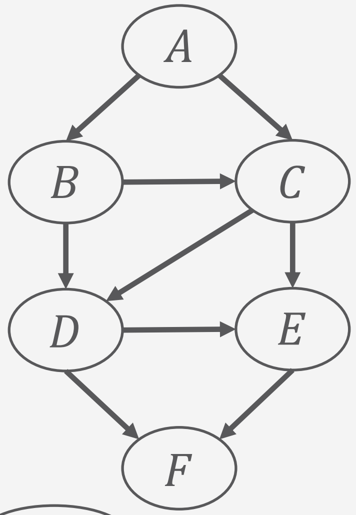


Umwandlung von MNs in BNs fügt in der Regel wesentlich mehr Kanten ins BN ein und führt damit zum Verlust von Unabhängigkeiten wie beim Moralisieren von BNs

Umwandlung von MNs in BNs

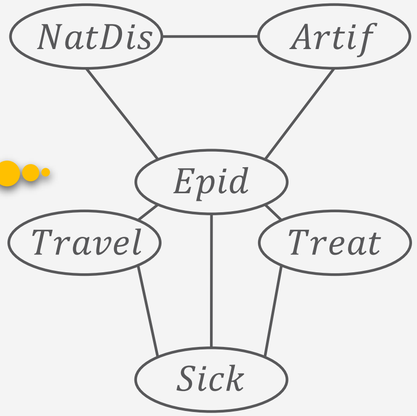
- Umwandlung nennt sich Triangulierung des MNs
  - Alle Zyklen sind in Dreiecke partitioniert
  - Ergebnis ist immer ein chordaler Graph
    - Chordal bzw. trianguliert: Alle Zyklen mit mindestens vier Knoten haben eine Sehne
      - *Sehne*: Kante, die nicht Teil des Zyklus ist aber zwei Knoten des Zyklus verbindet
      - Auch hier: alle Zyklen sind in Dreiecke partitioniert
- Wenn ein MN chordal ist, dann gibt es ein BN mit den gleichen Abhängigkeiten





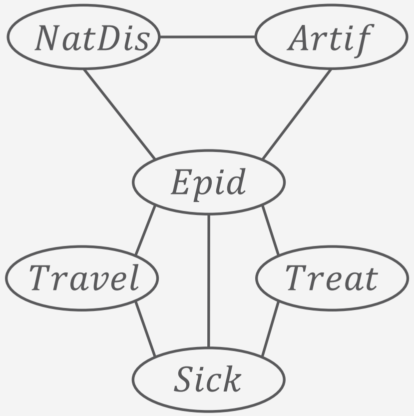
Es gilt sogar: MN und BN stellen die gleichen Abhängigkeiten dar *genau dann, wenn* der ungerichtete Graph chordal ist

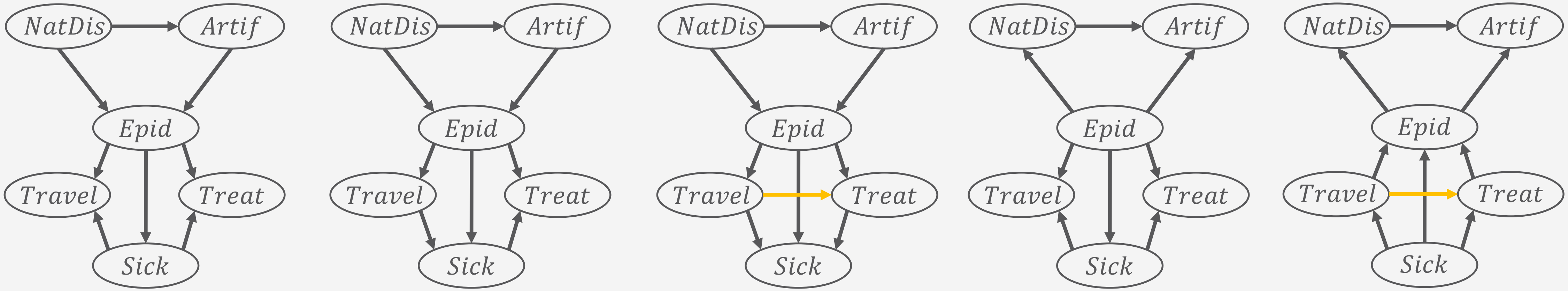
Wie sieht es hier aus?



Umwandlung von MNs in BNs

- Beispiel: Chordales MN
  - Ergebnisse bei einer Auswahl an unterschiedlichen Reihenfolgen
    - Reihenfolge nicht ganz beliebig, da auch mehr Kanten möglich; dann Unabhängigkeit verloren: $Treat \perp Travel \mid Epid, Sick$

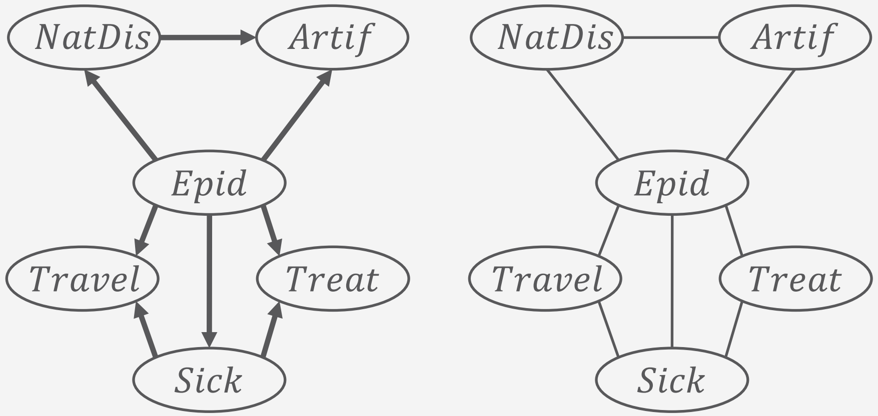


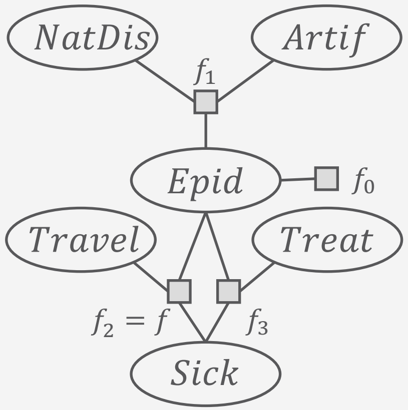


N, A, E, S, Tl, Tt $\quad$ N, A, E, Tl, S, Tt $\quad$ N, A, E, Tl, Tt, S $\quad$ E, N, A, S, Tl, Tt $\quad$ S, Tl, Tt, E, N, A

Umwandlung von Faktoren

- Unwandlung von chordalen MNs in BNs gegeben einer Reihenfolge der Knoten recht einfach
  - Dann sind die Unabhängigkeiten gleich
- Anzahl an CPDs kann allerdings zunehmen:
  - Beispiel: Im Faktormodell sind es *vier* Faktoren der Größen 2,8,8,8 für f_0, f_1, f_2, f_3
    - Das BN schreibt die *sechs* CPDs $P(E)$, $P(N \mid E)$, $P(A \mid N, E)$, $P(S \mid E)$, $P(Tl \mid E, S)$, $P(Tt \mid E, S)$ vor mit Größen 2,4,8,4,8,8
      - Wenn Boolesch, dann tatsächlich nur 1,2,4,2,4,4
- CPDs erhalten, indem man die entsprechenden Anfragen ans Faktormodell stellt → Thema 3: Exakte Inferenz
  - Anfragen: $P(E)$, $P(N \mid E)$, $P(A \mid N, E)$, $P(S \mid E)$, $P(Tl \mid E, S)$, $P(Tt \mid E, S)$





Zwischenzusammenfassung

- Paarweise, lokale und globale Unabhängigkeiten bzw. Markov Eigenschaften
  - Äquivalent in strikt positiven Verteilungen
- v-Strukturen in BNs
  - Bewirken, dass das Markov Blanket anders definitiert werden muss
  - Bewirken, dass Unabhängigkeiten verloren gehen können, wenn man ein BN in ein Faktormodell umwandelt
- Hammersley-Clifford Theorem: Eine Faktorisierung gibt es genau dann, wenn eine der Markov Eigenschaften gilt
- Kurz: Von Verteilungen zu Faktorisierung mit Hilfe von Unabhängigkeiten
- Umwandlung von MNs in BNs mittels Triangulierung → Ergebnis ein chordales BN
- Äquivalente Kodierung von Unabhängigkeiten zwischen BNs und MNs
  - Wenn keine Kanten bei der Umwandlung hinzugefügt werden
  - BN ein moraler Graph, dann MN mit gleichen Abhängigkeiten
  - Ungerichtete Graph ein chordaler Graph, dann kodieren BN und MN die gleichen Abhängigkeiten

Überblick: 2. Episodische PGMs

A. Probabilistische Modellierung
- Zufallsvariablen, vollständige gemeinsame Wahrscheinlichkeitsverteilung, Speicherkomplexität
- Inferenzaufgaben, Komplexität

B. Gerichtete Modelle: Bayes Netze (BNs)
- (Bedingte) Unabhängigkeiten, Faktorisierung
- Syntax, Semantik, graphische Darstellung, Speicherkomplexität

C. Ungerichtete Modelle: Faktormodelle
- Syntax, Semantik, graphische Darstellungen und deren Unterschiede, Speicherkomplexität
- Umwandlung von BNs zu Faktormodellen und deren graphischen Darstellungen

D. Unabhängigkeiten in PGMs
- Lokale, globale und paarweise Unabhängigkeiten
- Äquivalenzbedingungen von Faktormodellen und BNs

→ Exakte Inferenz in episodischen PGMs

Einordnung der Vorlesung: *Modell- und nutzenbasierter Agent*

- Nachfolgende Themen der Vorlesung
  2. Episodische PGMs
  3. Exakte Inferenz in episodischen PGMs
  4. Approximative Inferenz in episodischen PGMs
  5. Lernalgorithmen für episodische PGMs
  6. Sequentielle PGMs und Inferenz
  7. Entscheidungstheoretische PGMs