Bayesnetzmodelle (BNM)
in der Kardiologie

Vorgehensmodell - Ergebnisse

Claus Möbus - Heiko Seebold

Jan-Ole Janssen, Andreas Lüdtke,
Iris Najman, Heinz-Jürgen Thole

Besonderen Dank an:

Herrn Reinke (Münster)
und Herrn Dr. Rulands (Aachen)

Vorgehensmodell

Claus Möbus

Jan-Ole Janssen, Andreas Lüdtke,
Iris Najman, Heiko Seebold, Heinz-Jürgen Thole

(*neues) Vorgehensmodell:
qualitatives Modell (1/2)

- Nennung aller Variablen (Entitäten)
  - Expositionen, Krankheiten, Syndrome, Symptome, Alternativerklärungen, Folgekrankheiten, ..., etc
  - Problem der vorzeitigen Festlegung auf konkrete Begriffe
  - *Problem der Generierung neuer Variablen bzw. Ausprägungen (zB. latenter Variablen, Faktoren, Folgen, etc)
- *Erhebung „kausal vor-“ bzw. „kausal nach-“Relation
  - reflexive
  - transitive
  - Antisymmetrische
- *Pruefung der Dreiecksgestalt der Relationentabelle
- *Berechnung des Hassediagramms der Relation

(*neues) Vorgehensmodell:
qualitatives Modell

- *Hassediagramm ist minimaler DAG des BN
- *Prüfung der Berechtigung der Minimalität auf Grund der „Markov Blanket (MB)“-Eigenschaft
  - 1. Richtung: X_i => MB => Restnetz
  - 2. Richtung: X_i <= MB <= Restnetz
- bei Verletzung der MB-Eigenschaft Einzug neuer Kanten

Vorgehensmodell: quantitatives Modell

- **Erhebung probabilistischer Regeln**
  - in „Kausal-“richtung
  - in diagnostischer Richtung
  - in gemischter Richtung

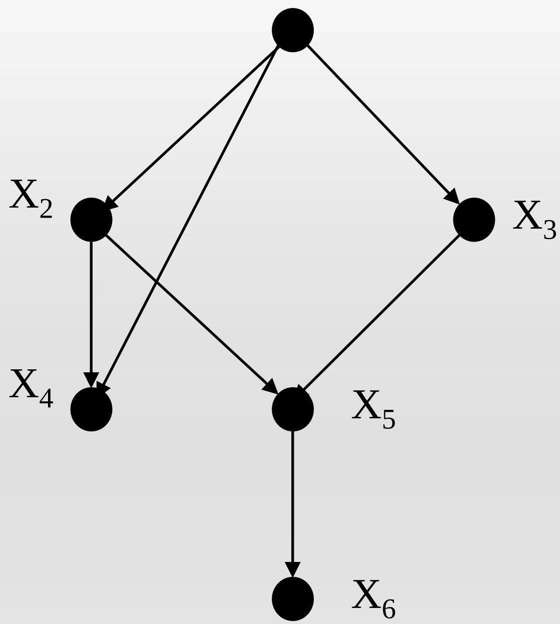
- **Schätzung des Restmodells nach dem Prinzip der maximalen Entropie (ME)**

Definition: Bayes Netz

- set of *variables* X = {X_1, ... , X_n}
- set of *directed edges* between variables
- each variable has a finite set of *mutually exclusive states*
- variables together with the directed edges form a *DAG*
- to each variable Y with parents X_1, ... , X_n, there is attached the *conditional potential table* P(Y| X_1, ... , X_n)
- the joint probability distribution P(X) is the *product of all potentials* specified in BN

$$P(X) = \prod_i P(X_i \mid parents(X_i))$$

Gemeinsame Verteilung als Produkt bedingter Verteilungen



$$P(X_6, X_5, X_4, X_3, X_2, X_1)$$
$$= P(X_6 | X_5) P(X_5 | X_3, X_2) P(X_4 | X_2, X_1) P(X_3 | X_1) P(X_2 | X_1) P(X_1)$$
$$= \prod_i P(X_i | \Pi_{X_i})$$

wobei : Π_{X_i} = direkte Vorgänger von X_i

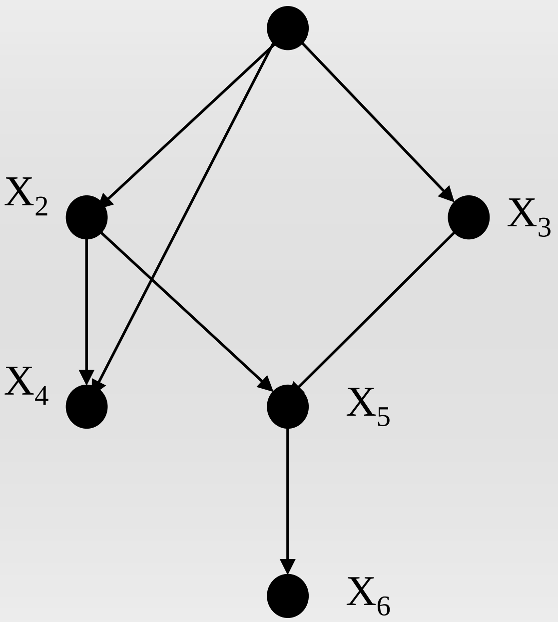
Definition: Markov Blanket (MB)

JENSEN, 2001

- **set of *parents* of Y = {X_1, ... , X_m}**
- **set of *children* of Y = {Z_1, ... , Z_n}**
- **set of *lovers* of Y = {U_1, ... , U_p}, sharing a child with Y**
- **If all variables in the MB for Y are instantiated, the Y is d-separated from the rest of the network**

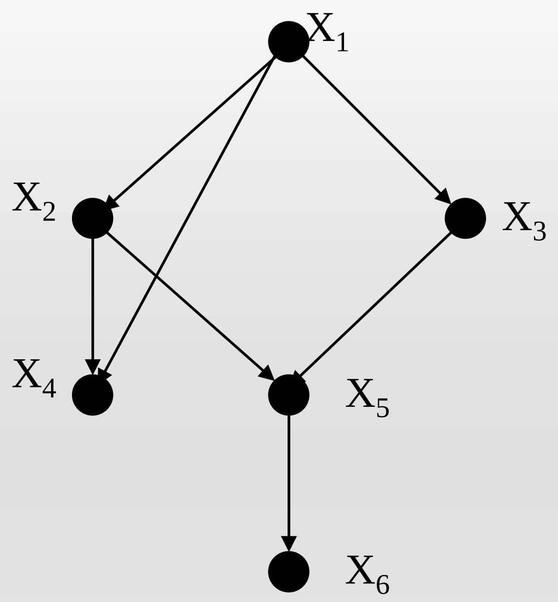
$$\{X \perp N - MB \mid MB\}$$

Markov Blankets



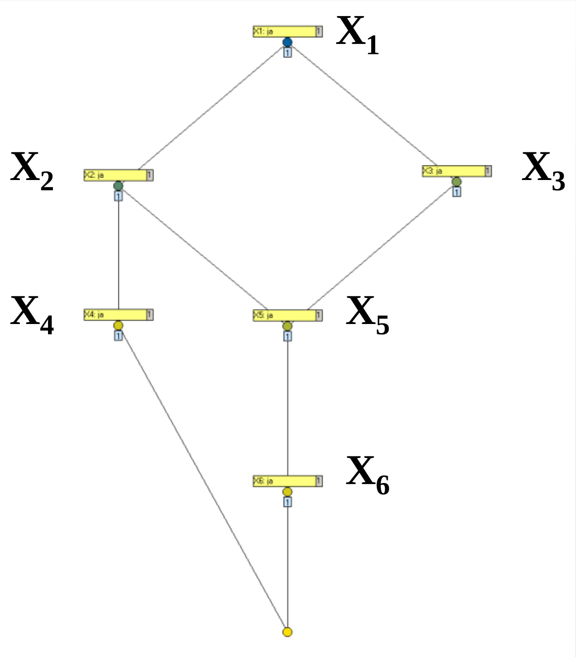
MB(X_i)	parents	children	lover
X1		X2,X3,X4	X2
X2	X1	X4,X5	X1,X3
X3	X1	X5	X2
X4	X1,X2		
X5	X2,X3	X6	
X6	X5		

„wahres“ BN und „Kausal-vor-“Relation



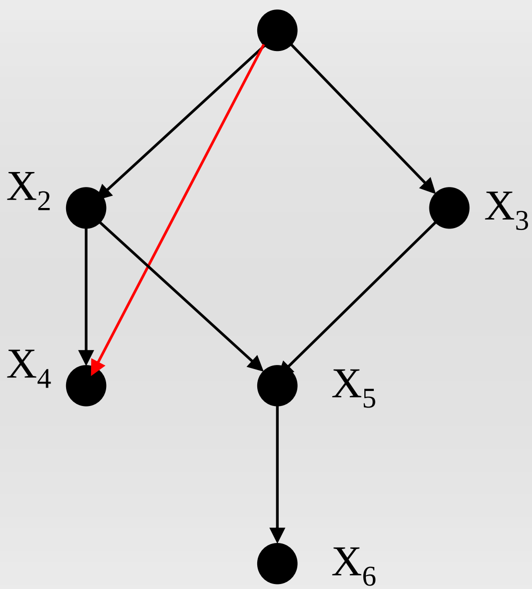
Objekt 1 von 6 Eigenschaft 1 von 6	X1 kausal_vor	X2 kausal_vor	X3 kausal_vor	X4 kausal_vor	X5 kausal_vor	X6 kausal_vor
X1	ja					
X2	ja	ja				
X3	ja	nein	ja			
X4	ja	ja	nein	ja		
X5	ja	ja	ja	nein	ja	
X6	ja	ja	ja	nein	ja	ja

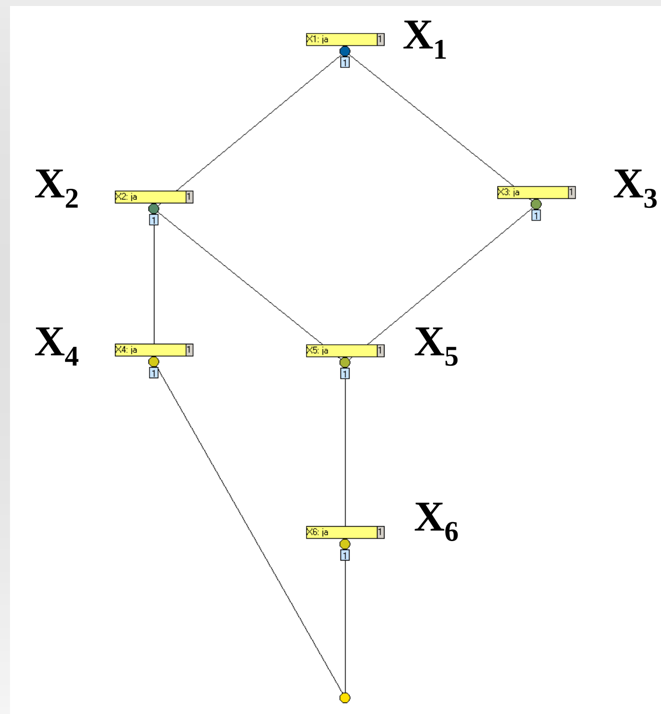
aus „Daten" konstruiertes Hassediagramm



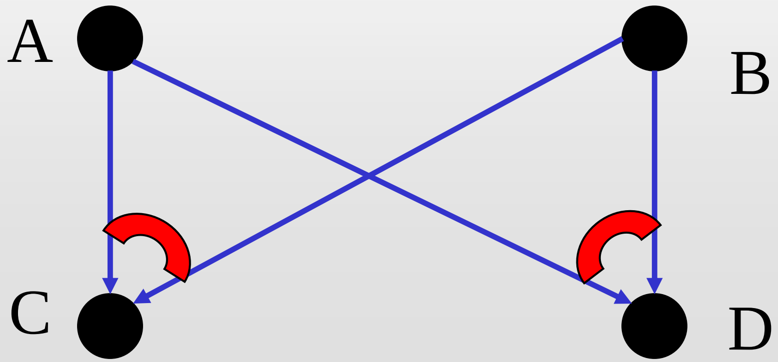
Objekt 1 von 6 Eigenschaft 1 von 6	X1 kausal_vor	X2 kausal_vor	X3 kausal_vor	X4 kausal_vor	X5 kausal_vor	X6 kausal_vor
X1	ja					
X2	ja	ja				
X3	ja	nein	ja			
X4	ja	ja	nein	ja		
X5	ja	ja	ja	nein	ja	
X6	ja	ja	ja	nein	ja	ja

„wahres“ BN und Begriffsverband der Variablen





naives BNM
4-Knoten-Problem



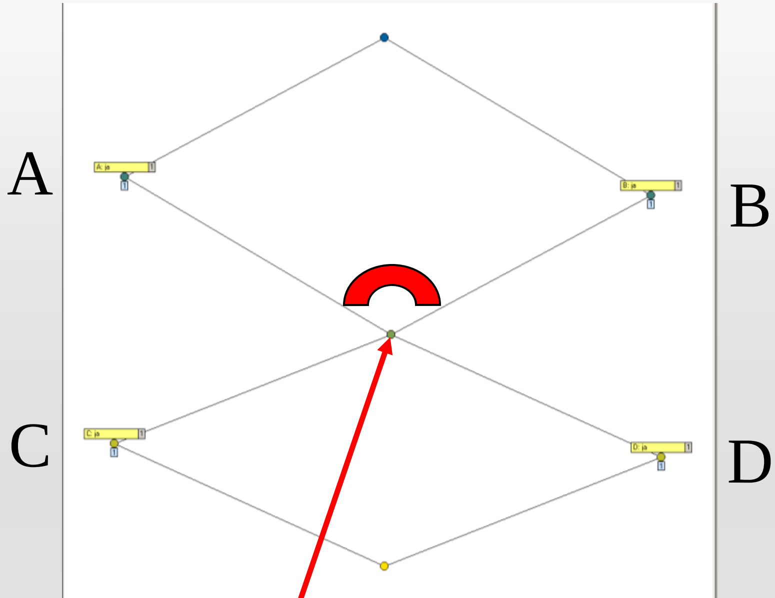
BN mit 2+2+8+8=20 Parms;

4 explizite Knoten

Objekt 4 von 4 igenschaft 2 von	A kausal vor	B kausal vor	C kausal vor	D kausal vor
A	ja			
B		ja		
C	ja	ja	ja	
D	ja	ja		ja

minimales BN durch Hassediagramm

4-Knoten-Problem



BN mit 2+2+8+2+2=16 Parms;

4 explizite Knoten;

1 neuer latenter Knoten

Objekt 4 von 4 igenschaft 2 von	A kausal vor	B kausal vor	C kausal vor	D kausal vor
A	ja			
B		ja		
C	ja	ja	ja	
D	ja	ja		ja

Vorgehensmodell:
quantitatives Modell

- **Erhebung probabilistischer Regeln**
  - in „Kausal-“richtung
  - in diagnostischer Richtung
  - in gemischter Richtung

- **Schätzung des Restmodells nach dem Prinzip der maximalen Entropie (ME)**

Bedingte Schätzung des Modells
Prinzip der maximalen Entropie (ME)

- **Menge R probabilistischer Regeln**

$$R = \{A_1 \rightarrow B_1[x_1], \dots, A_n \rightarrow B_n[x_n]\}$$

- **erfüllende Verteilung P**

$$P \models A \rightarrow B[x] \;\; \text{gdw.} \;\; P(B \mid A) = x \quad \text{mit} \;\; P(A) > 0$$

- **bedingte Schätzung P* der Verteilung P unter ME**

$$\max_{P \models R} H(P) = -\sum_{\omega} P(\omega) \cdot \log_2 P(\omega)$$

- **Eindeutig lösbares Optimierungsproblem (Csiszár, 1975)**

$$P^* = ME(R)$$

Ergebnisse

Heiko Seebold

Jan-Ole Janssen, Andreas Lüdtke,
Claus Möbus, Iris Najman, Heinz-Jürgen Thole

Besonderen Dank an:

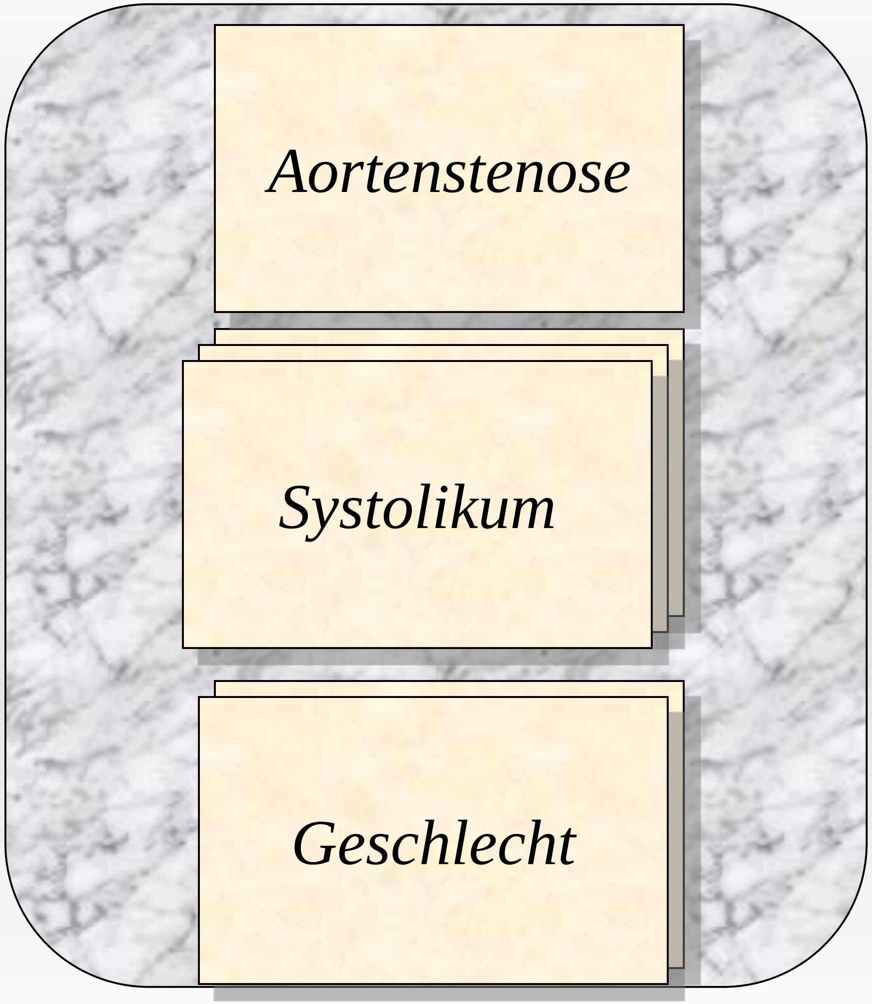
Herrn Reinke (Münster)
und Herrn Dr. Rulands (Aachen)

Modellierung einer Krankheit

- **Mediziner :**
  - **Herr Reinke (Münster),**
  - **Herr Dr. Rulands (Aachen)**
- **Modellierte Krankheit : Aortenstenose**
- **Modellierung wurde getrennt durchgeführt**

Erhebung und Ordnung der Variablen

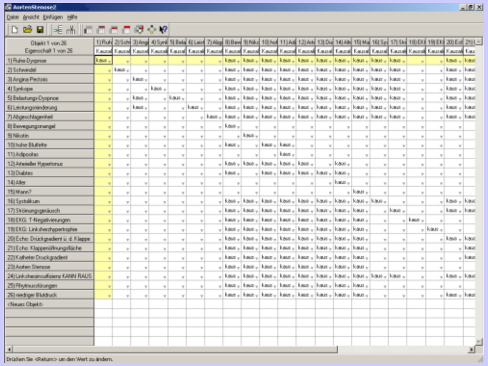
- **Variablen können Symptome, Expositionen, Krankheiten, etc. sein.**
- **Pro Netz ca. 30 verschiedene Variablen erhoben**
- **Iteratives Betrachten einzelner Variablen**
- **Ordnung der restlichen Variablen mittels der Relation „kausal vor“**



Modellierung des qualitativen Netzes

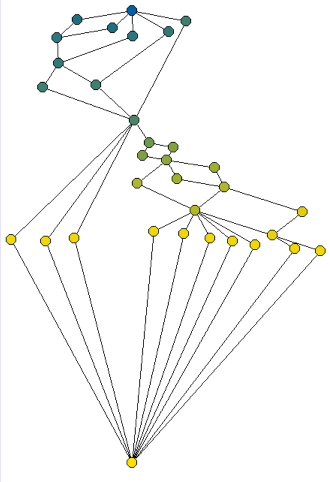
1. Schritt

- Eingabe der Relationen in eine Tabelle
- Prüfung der Dreiecksgestalt



2. Schritt

- Erzeugung des minimalen Graphen (Hasse-Diagramm)
- ➡ Revision durch Unabhängigkeitsdialog notwendig



3. Schritt

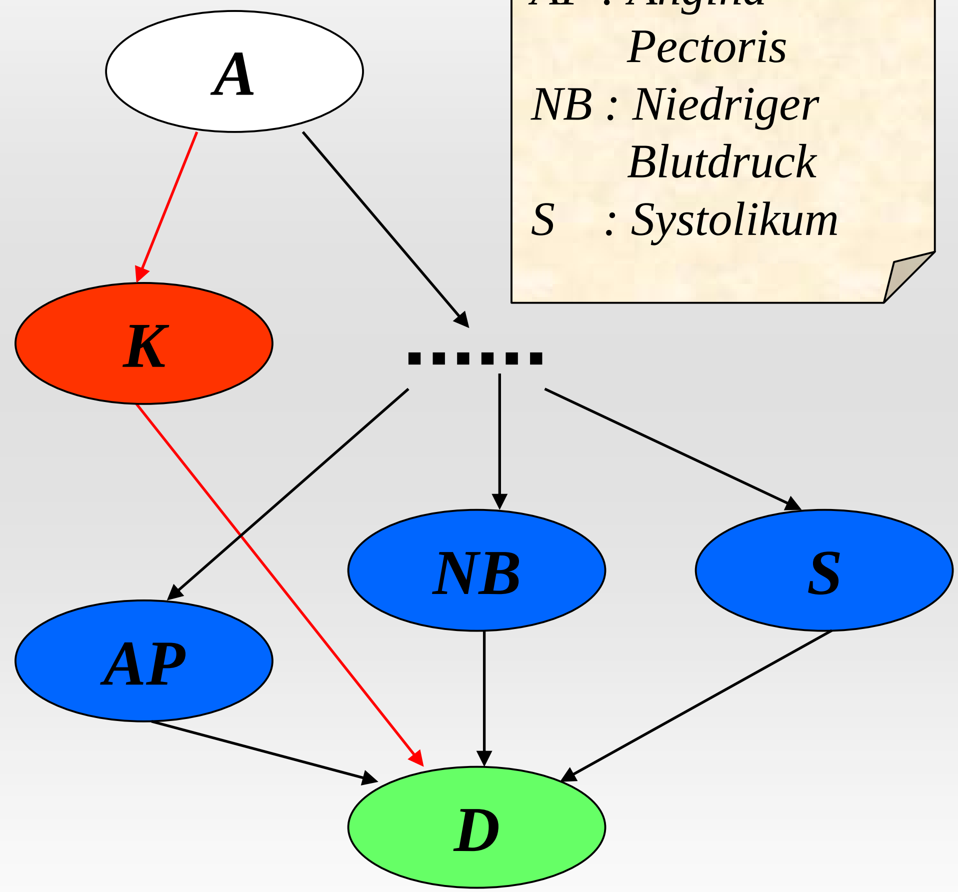
- Übertragung in ein Modellierungs-werkzeug für Bayes-Netze nach praktikablen Gesichtpunkten
- Entwickelt von unserer Abteilung
- Unterstützt den Entropieansatz zur Berechnung fehlender Wahrscheinlichkeiten

Medikus

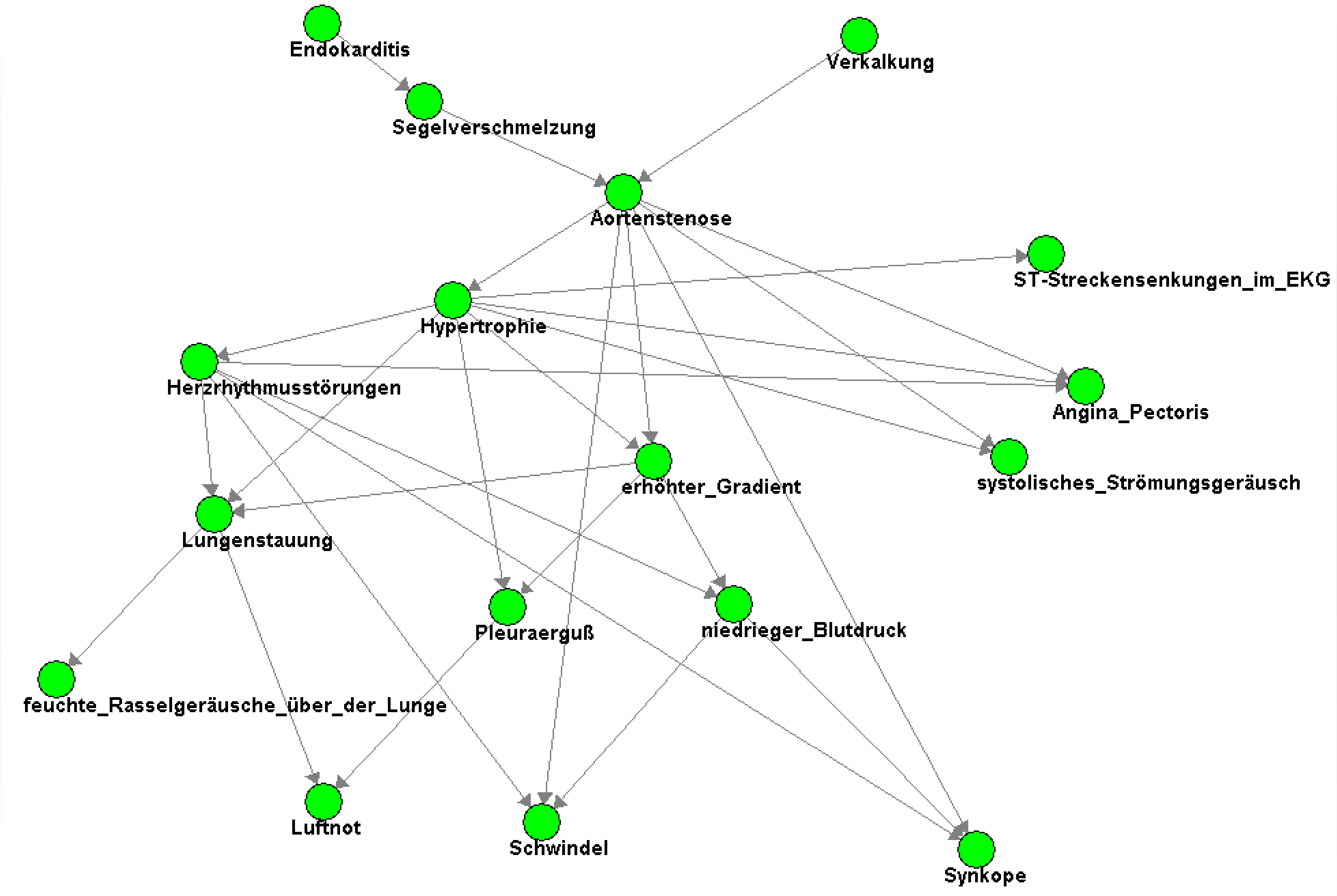
Unabhängigkeitsdialog

- **„Stellen Sie sich vor, dass** bekannt ist**, ob** Angina Pectoris, Niedriger Blutdruck **und** Systolikum **vorliegen. Gibt es andere Variablen im Netz, die zusätzliche Informationen zur Wahrscheinlichkeit von** Dyspnoe **liefern?“**
- **Antwort „Ja“ führt zur Modellrevision**
- **Neuer Knoten** Kontraktilität (K)

A : Aortenstenose
AP : Angina Pectoris
NB : Niedriger Blutdruck
S : Systolikum



Bayesnetz zur Aortenstenose (Ausschnitt)



Erhebung der Wahrscheinlichkeiten

- Problematisch, da oft nicht alle Wahrscheinlichkeiten bekannt
- Kausal Wahrscheinlichkeiten schwieriger anzugeben als diagnostische
- Leichter Umgangssprachliche Aussagen zu treffen, als konkrete Zahlen zu nennen
- Problem: Durchschnittliche Werte
- Berechnung fehlender Wahrscheinlichkeiten nach dem Prinzip der maximalen Entropie

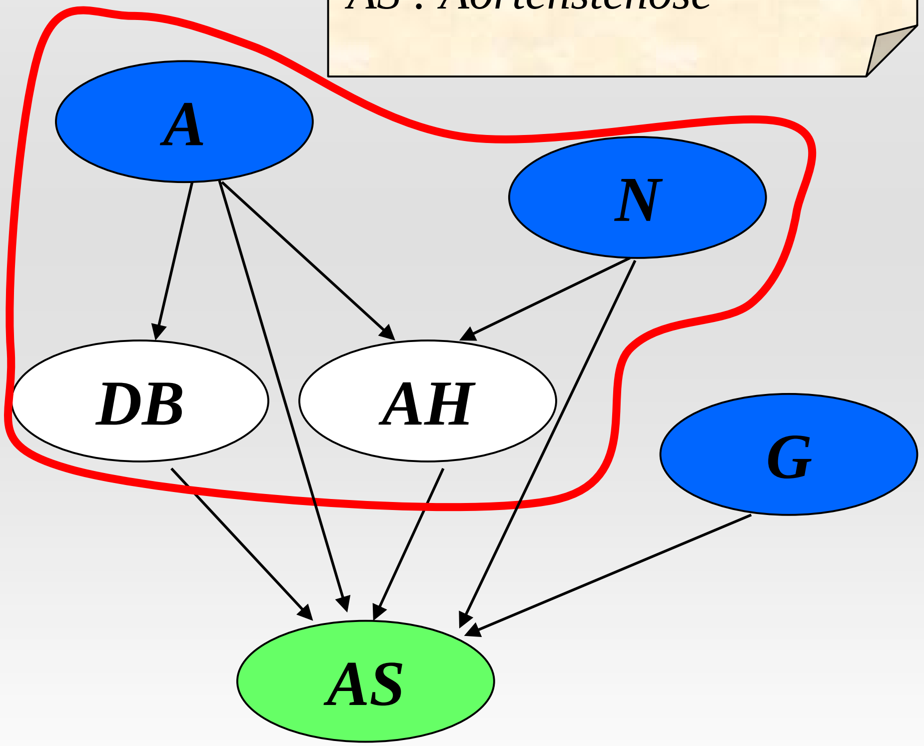
Linguistischer Term	Gültigkeitsbereich
fast sicher	0.99 - 0.91
…	…
Fünfzig zu fünfzig	0.54 – 0.46
…	…
unwahrscheinlich	0.27 – 0.19
…	…

Lichtenstein, Newman - 1967

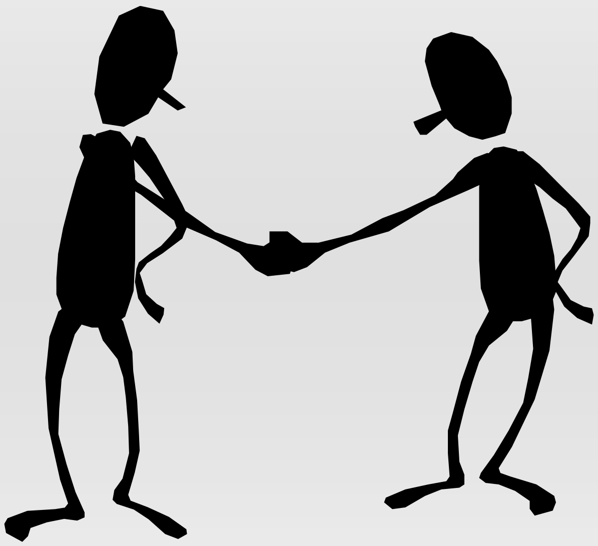
Erhebung der Wahrscheinlichkeiten

- Für P(A), P(N) und P(G) statistische Daten notwendig
- Kontext beachten
- P(AS| G, A, N, DB, AH) für alle Kombinationen schwierig anzugeben
- Leichter: Einteilung der Auslöser in Risikofaktoren (RF)
- P(AS = leicht | RF vorhanden >= 3, G = männlich) = 0.28

A : Alter
N : Nikotin
AH : Arterieller Hypertonus
DB : Diabetis
G : Geschlecht
AS : Aortenstenose



- 2 kausale Netze erstellt
- Ca. 30 Variablen pro Netz
- Ca. 150 Wahrscheinlichkeiten erhoben



Vielen Dank für Ihre Aufmerksamkeit ...