OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

INF FAKULTÄT FÜR INFORMATIK

Bayesian Networks: Theory and Applications

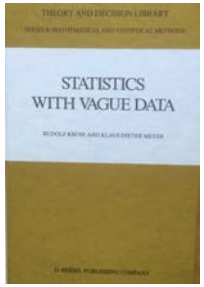
Prof. Dr. Rudolf Kruse

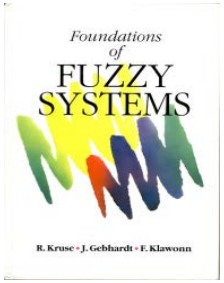
Vrije Universiteit Brussel

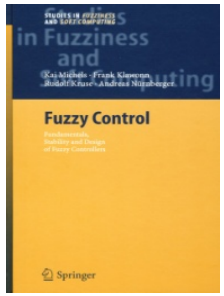
22.4.2022

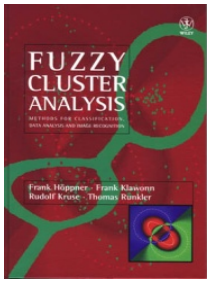
Computational Intelligence Group Magdeburg

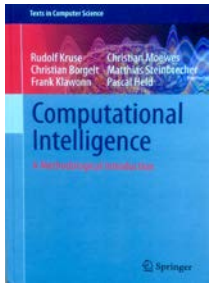
Research: Data Sciences, Intelligent Systems

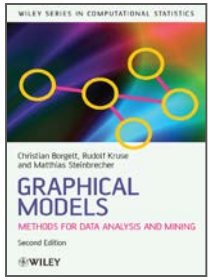




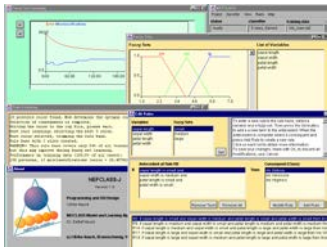








Software Tools: NEFCLASS, Information Miner,…



Transfers: Industrial Projects (BT, SAP, Siemens, Volkswagen, …), Spin Offs

Property planning - Volkswagen

Property family	Car body	Motor	Radio	Doors	Seat cover	Makeup mirror	…
Property	Hatch-back	2.8 L 150 kW Otto	Type alpha	4	Leather, Type L3	yes	…

Complexity

- About 200 variables
- Typically 4 to 8, but up to 150 possible instances per variable
- More than 2^{200} possible combinations available

Knowledge about the Planning

Rules

- 10000 Technical Rules for Item Combinations, e.g.

$$\text{IF Motor} = m_4 \text{ AND Heating} = h_1$$

$$\text{THEN Generator} \in \{g_3, g_4, g_5\}$$

- Often 6-dimensional, sometimes more than 10 dimensions

- 500000 marketing oriented rules (with uncertainty)

- The rules are often changing

Data

- Specification of millions of built cars

Planning Tasks

Calculation of part demands

Compute the installation rate of a given item combination

Simulation

Analyze customers‘ preferences with respect to those persons who use a navigation system in a VW Polo

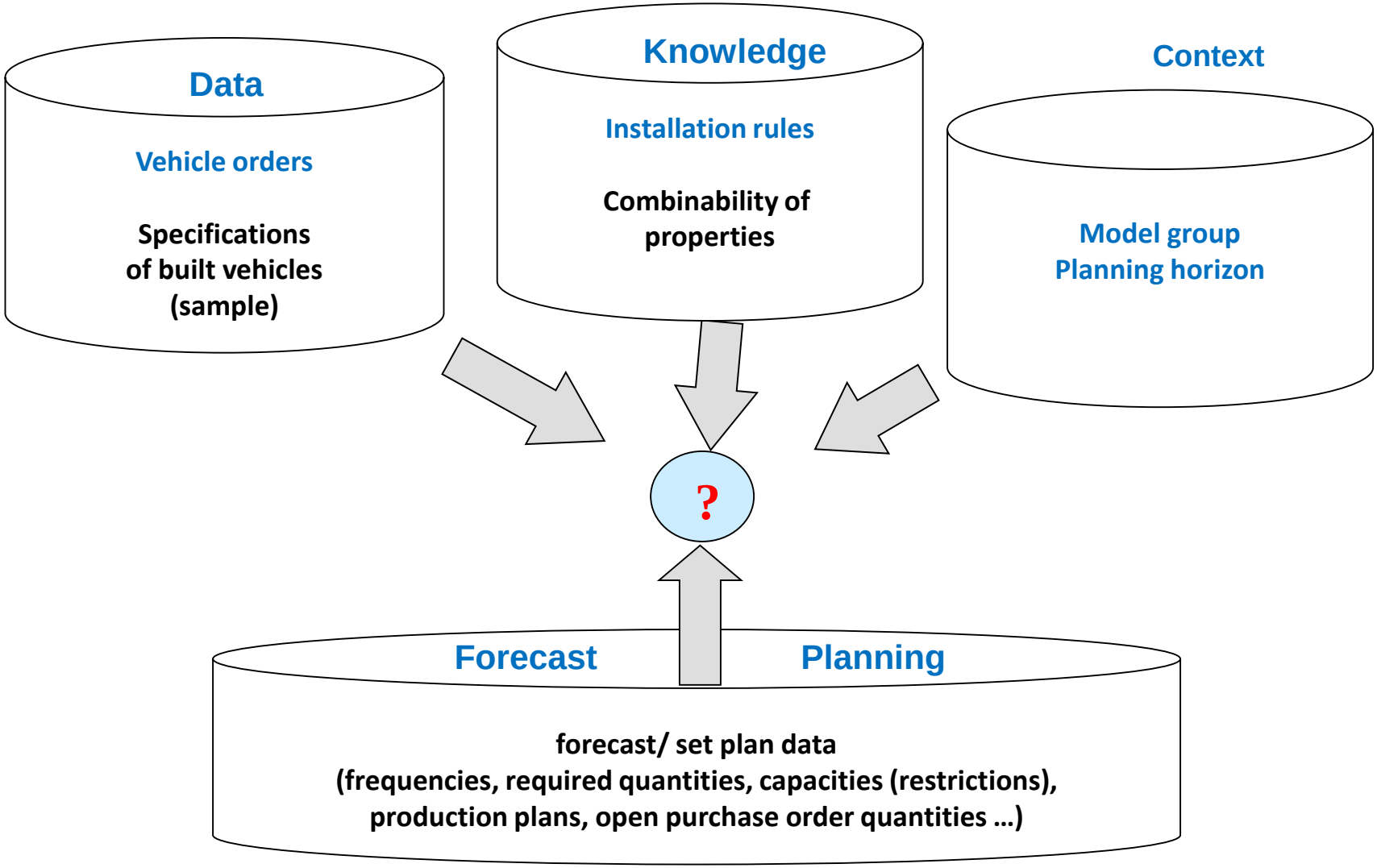
Marketing and Sales stipulation

Installation rate of Navigation system increase from 20% to 30%

Capacity Restrictions

Maximum availability of seat coverings in leather is 5000

How to manage the uncertain information about planning?



How to handle uncertain Information in AI Systems?

Method 1 : Rule Based Approaches

Example Mycin (1975) – Expert System for diagnosing bacterial infections

500 (uncertain) decision rules as knowledge base

If 1) the gram stain of the organism is gramneg, and
2) the morphology of the organism is rod, and
3) the aerobicity of the organism is anaerobic
then there is suggestive evidence (0.6) that the identity of the organism is bacteroides

Certainty Factor cf: $-1 \leq cf \leq 1$

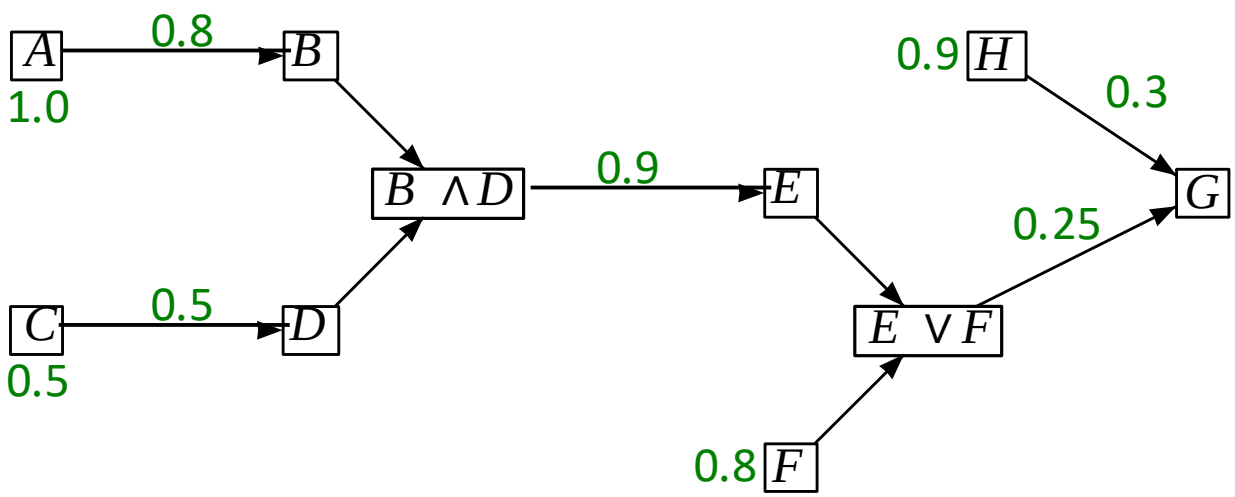
Rules

$A \rightarrow B$ [0.80]
$C \rightarrow D$ [0.50]
$B \wedge D \rightarrow E$ [0.90]
$E \vee F \rightarrow G$ [0.25]
$H \rightarrow G$ [0.30]

Facts

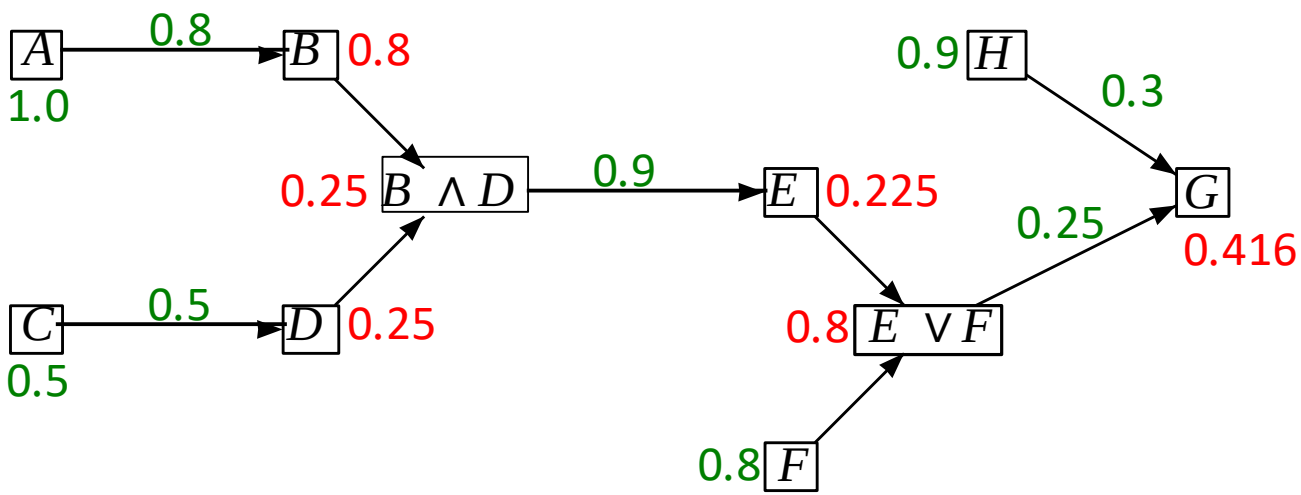
A [1.00]
C [0.50]
F [0.80]
H [0.90]

Tree



Rules for conjunction disjunction, serial und parallel combination combination, e.g.

CF(B ∧ D) = min{CF(A), CF(B)} , CF(D) = CF(C) x CF(C→D), CF(G)= 0.27 + 0.2 − 0.27 · 0.2 = 0.416



Drawback of this method:

- No modelling of (in) dependences (CF ist truth functional !).
- Reasoning in one direction
- Manual Design of rules and certainty factors

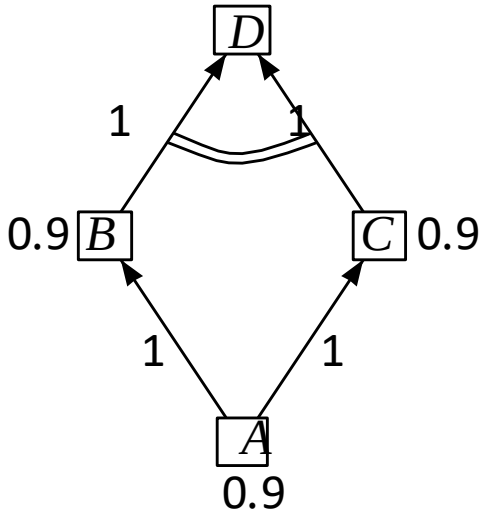
Opinion at that time:

- Probabilistic solutions are too complex for big systems

The CF rule combination scheme is inconsistent in general.

Example: CF(A) = 0.9, CF(D) =?

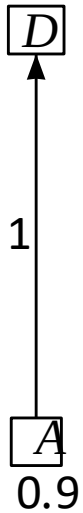
CF(D) = 0.9 + 0.9 − 0.9 · 0.9 = 0.99



A→B
A→C
B→ D
C→ D

CF(D) = 0.9

vs.



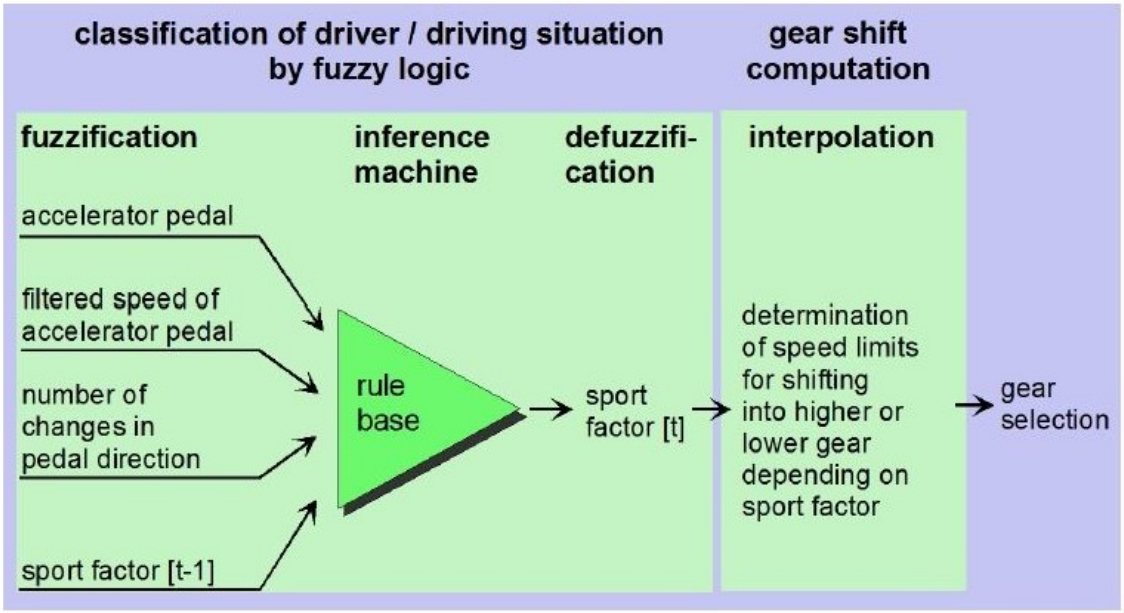
A→D

Certainty factor is increased just because (the same) evidence is transferred over two different (parallel) paths!

Example Fuzzy Control

Automatic Gear Box for VW Beetle (1995), in series line

Seven Fuzzy rules evaluate the sportiveness of driver 12 times per second

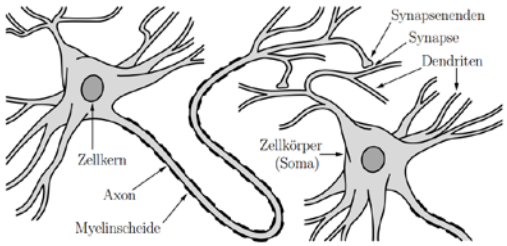


The rules were found by fuzzy clustering.

How to find consistent rules in big applications?

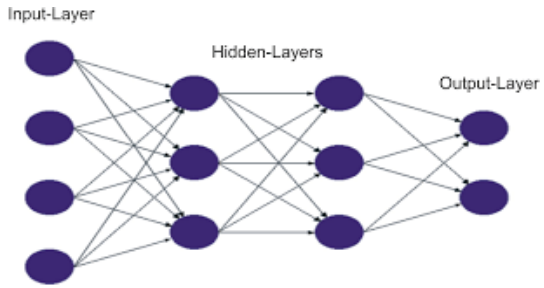
Method 2 Data Driven Approaches: Big Data Machine Learning, Deep Learning

Human Brain

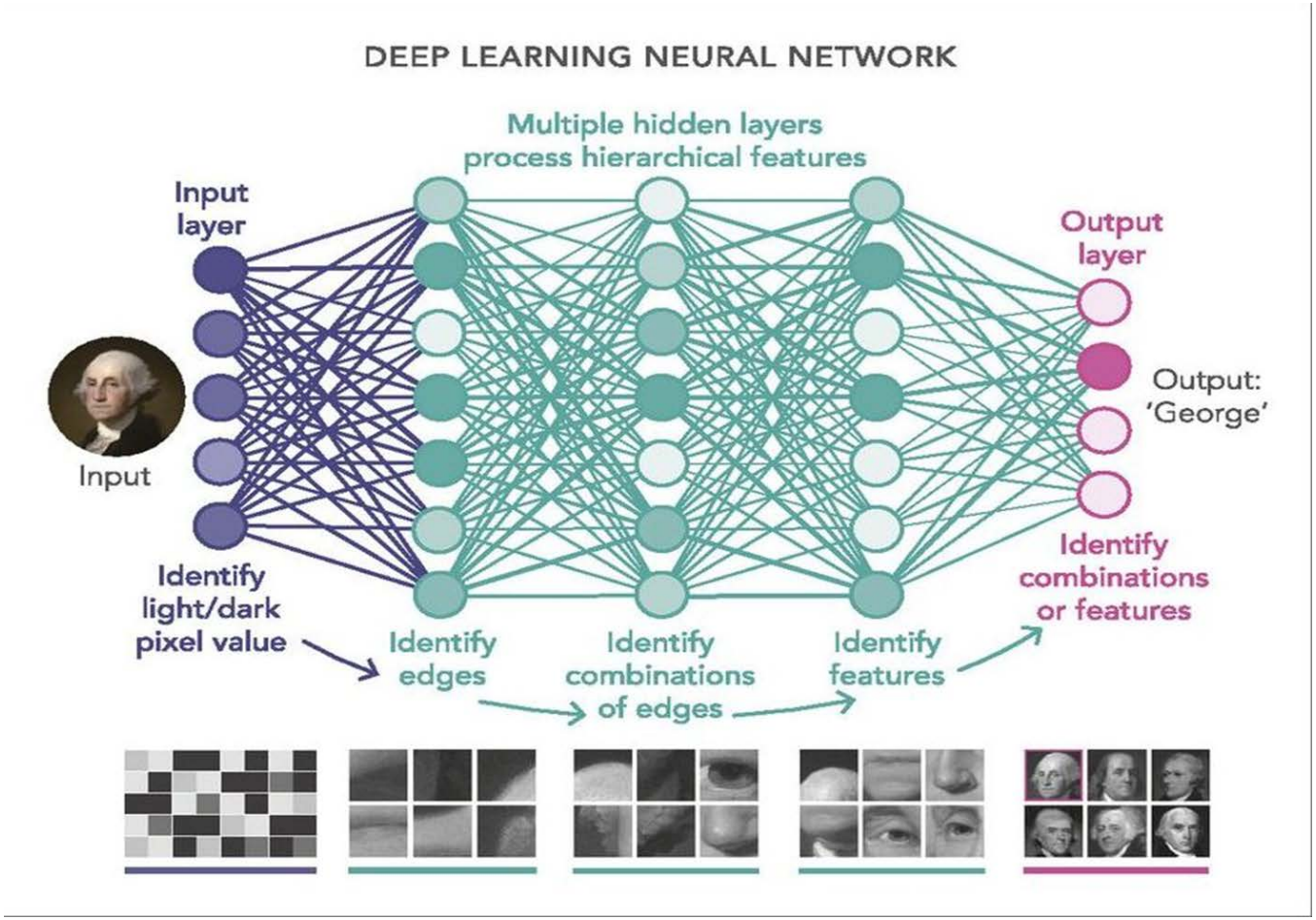


Approx. 86 000 000 000 Neurons
1000-10000 inputs per neuron

Nature Inspired Models



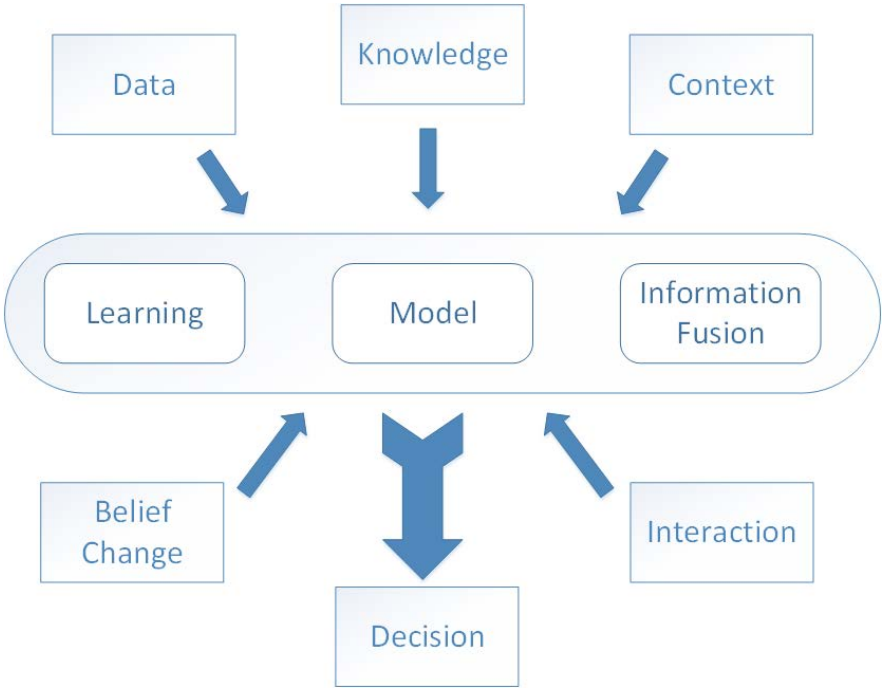
Artificial Neuronal Networks
Learning Back-Propagation
Deep Learning



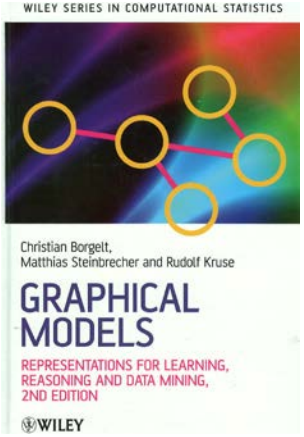
- Lots of data, high performance computing, clever algorithms are needed
- Typical Problems with pure data driven approaches: explainability, robustness, safety

Method 3 Hybrid Systems (Data + Knowledge)

Vision: A hybrid system accesses many sources and data, merges this information, filters and evaluates this information from the respective context, interacts with other systems and users, learns from information, uses knowledge about causalities, dependencies, associations, hits these base decisions, etc.



Our hybrid modeling approach for property planning



Installation Rates as Subjective Probabilities:

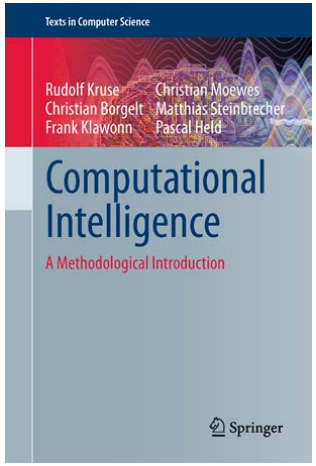
High dimensional finite probability space

Handling Complexity by using Decomposition:

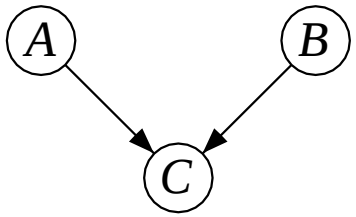
Instead of one high dimensional global model use several connected local low dimensional models

Implementation with Graph Based Probabilistic Models:

Relational, Bayes, and Markov Networks



How to find a suitable decomposition of a probability space?



A,B,C (Random) Variables

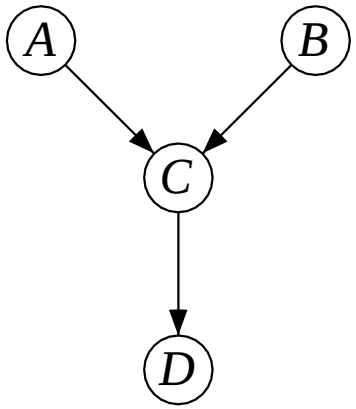
A quality of ingredients

B cook‘s skill

C meal quality

If C is not known, A and B are independent.

If C is known, then A and B become (conditionally) dependent given C.



A,B,C,D (Random) Variables

A quality of ingredients

B cook‘s skill

C meal quality

D restaurant success

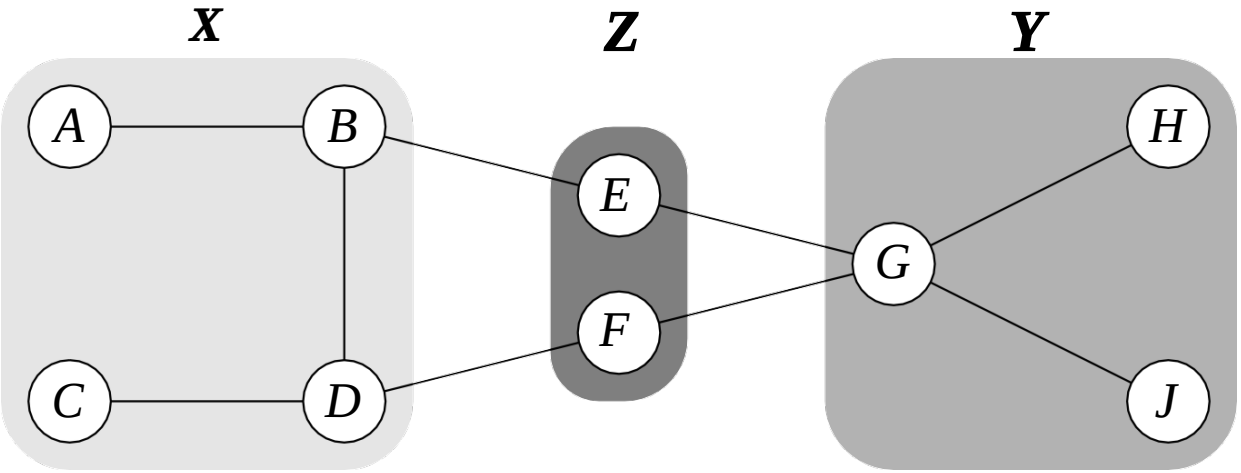
If nothing is known about the restaurant success or meal quality or both, the cook’s skills and quality of the ingredients are unrelated, that is, independent.

However, if we observe that the restaurant has no success, we can infer that the meal quality might be bad.

If we further learn that the ingredients quality is high, we will conclude that the cook’s skills must be low, thus rendering both variables dependent.

Decomposition: P(A,B,C,D) = P(A)P(B)P(CIA,B)P(DIC)

Separation in undirected graphs (u-Separation)



Z u-separates X from Y if every path from X to Y is blocked by a node in Z .

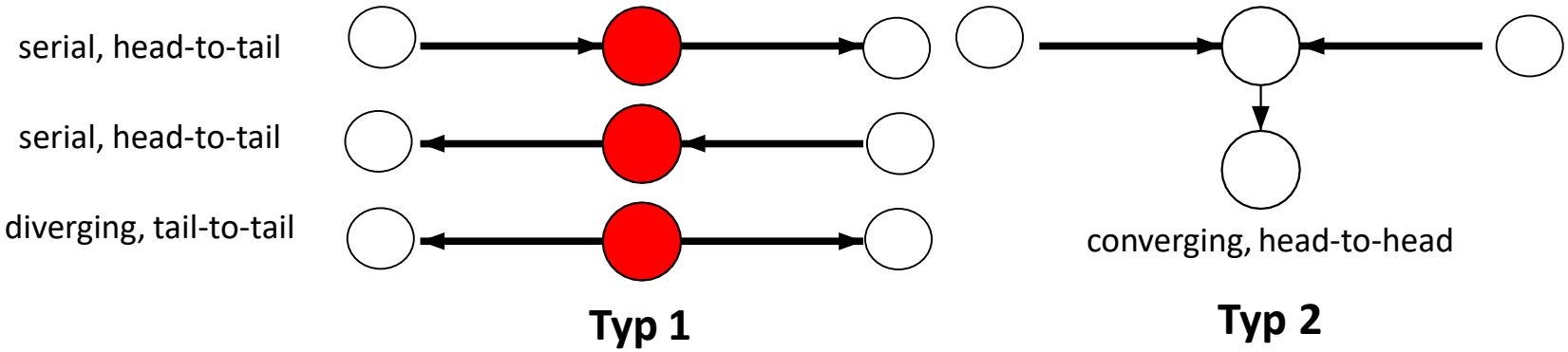
In directed graphs the decomposition is more tricky.

Separation in directed Graphs (d-separation)

Z d -separates **X** from **Y** if all paths (pathes in reverse to arrows are allowed) from **X** to **Y** are blocked by a node in **Z.**

A node *A* is blocking a path, if its edge directions along the path

- are of type 1 and $A \in$ ***Z,*** or
- are of type 2 und neither A nor one of its descendants is in **Z.**

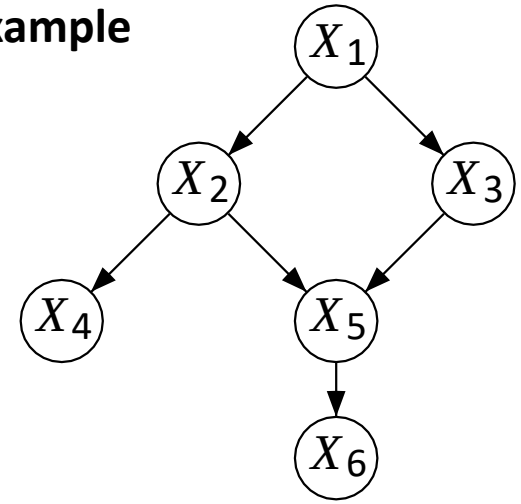


A **Bayesian network** is a directed acyclic graphs (DAG) of random variables **X**v with the property

$$p(x) = \prod_{v \in V} p\left(x_v \mid x_{\mathrm{pa}(v)}\right)$$

p(X) denotes the joint distribution of the random variables, pa(*v*) is the set of parents of *v* (i.e. those vertices pointing directly to *v* via a single edge).

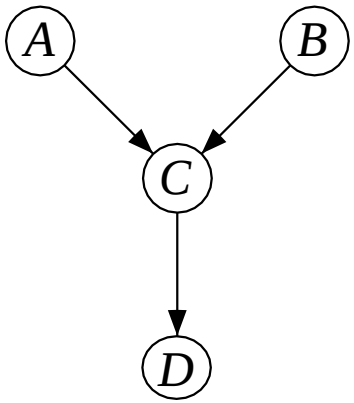
Example



$$\begin{aligned} P(X_1, \ldots, X_6) = {} & P(X_6 \mid X_5) \cdot \\ & P(X_5 \mid X_2, X_3) \cdot \\ & P(X_4 \mid X_2) \cdot \\ & P(X_3 \mid X_1) \cdot \\ & P(X_2 \mid X_1) \cdot \\ & P(X_1) \end{aligned}$$

Theorem If Z d-separates X und Y , then X and Y are conditionally independent with respect to Z

Qualitative Knowledge: DAG



Quantitative Knowledge: Conditional Probabilities

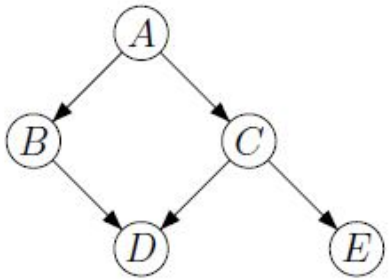
P(A,B,C,D) = P(A)xP(B)xP(CIA,B)xP(DIC)

Modelling with Bayesian Networks

A is a possible cause for B and also an explanation for C. Both in turn could explain that D is true. E may also be associated with D.

Variables	Values
A	a1, a2
B	b1, b2
C	c1, c2
D	d1, d2
E	e1, e2

DAG



Probabilities

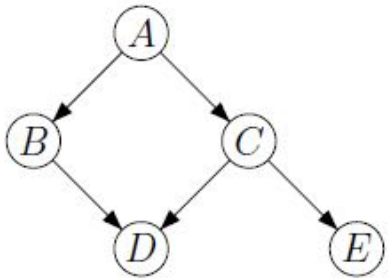
$$
\begin{aligned}
P(e_1 \mid c_1) &= 0.8 \\
P(e_1 \mid c_2) &= 0.6 \\
P(d_1 \mid b_1, c_1) &= 0.8 \\
P(d_1 \mid b_1, c_2) &= 0.8 \\
P(d_1 \mid b_2, c_1) &= 0.8 \\
P(d_1 \mid b_2, c_2) &= 0.05 \\
P(b_1 \mid a_1) &= 0.8 \\
P(b_1 \mid a_2) &= 0.2 \\
P(c_1 \mid a_1) &= 0.2 \\
P(c_1 \mid a_2) &= 0.05 \\
P(a_1) &= 0.2
\end{aligned}
$$

Reasoning with Bayesian Networks

E is observed . The „Belief“ in D changes.

Variables	Values
A	a1, a2
B	b1, b2
C	c1, c2
D	d1, d2
E	e1, e2

DAG



Probabilities

$$P(e_1 \mid c_1) = 0.8$$
$$P(e_1 \mid c_2) = 0.6$$

$$P(d_1 \mid b_1, c_1) = 0.8$$
$$P(d_1 \mid b_1, c_2) = 0.8$$
$$P(d_1 \mid b_2, c_1) = 0.8$$
$$P(d_1 \mid b_2, c_2) = 0.05$$

$$P(b_1 \mid a_1) = 0.8$$
$$P(b_1 \mid a_2) = 0.2$$

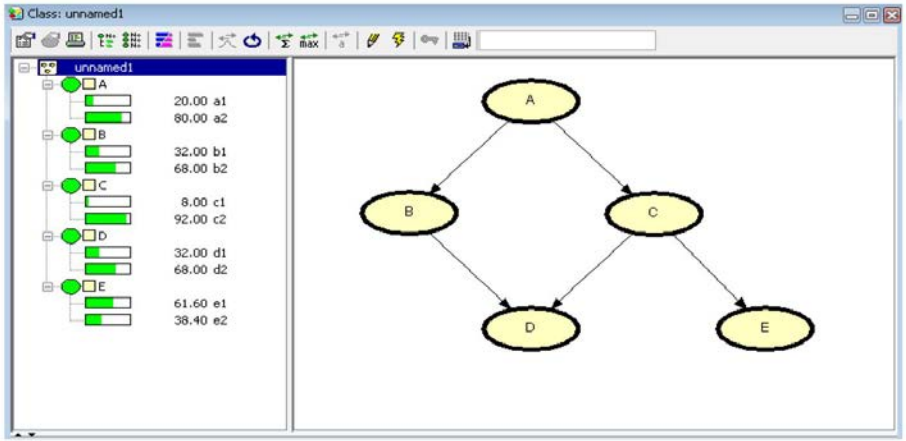
$$P(c_1 \mid a_1) = 0.2$$
$$P(c_1 \mid a_2) = 0.05$$

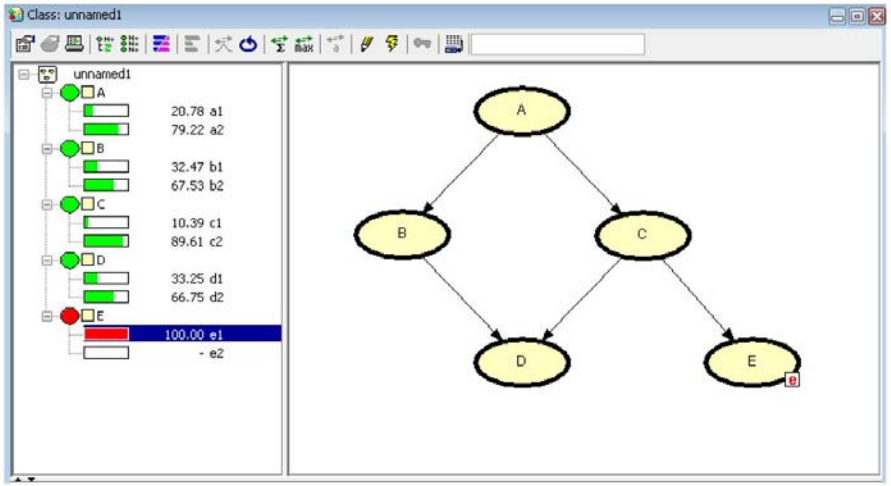
$$P(a_1) = 0.2$$

$$P(A, B, C, D\ E) = P(A)P(B \mid A)P(C \mid A)P(D \mid BC)P(E \mid C)$$

$P(d_1)=0.3200$, $P(d_1/e_1)=0,3325$

There are many tools for handling BN‘s.
Free Download of the state of the art BN-System Hugin Lite 8.9

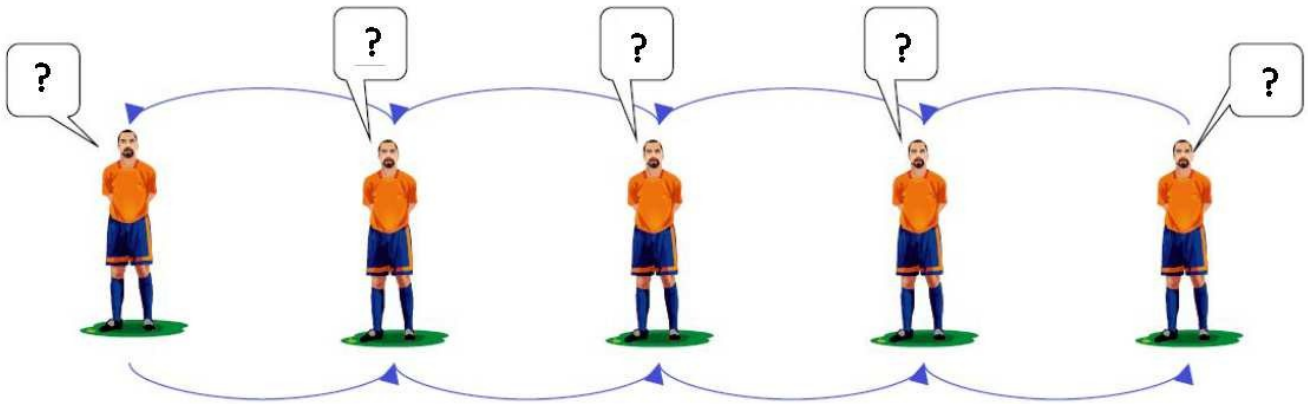




How to find efficient algorithms for large Bayesian Networks?

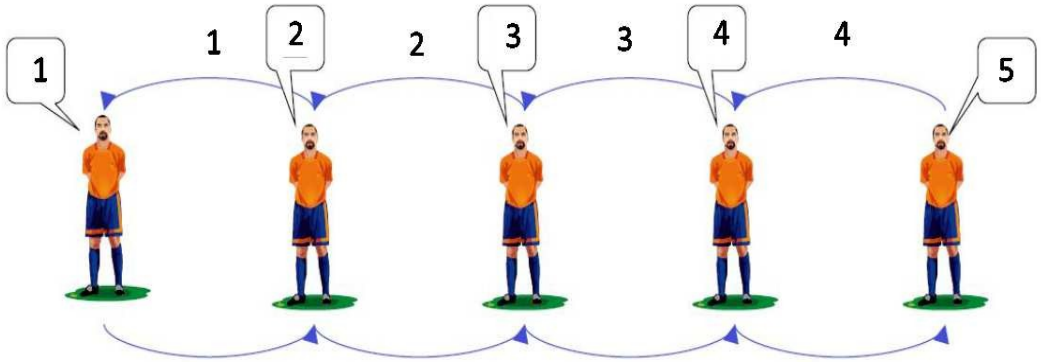
Solution: Global information can be shared locally by every entity.
New knowledge is distributed by message passing.

Example: How many people are we?

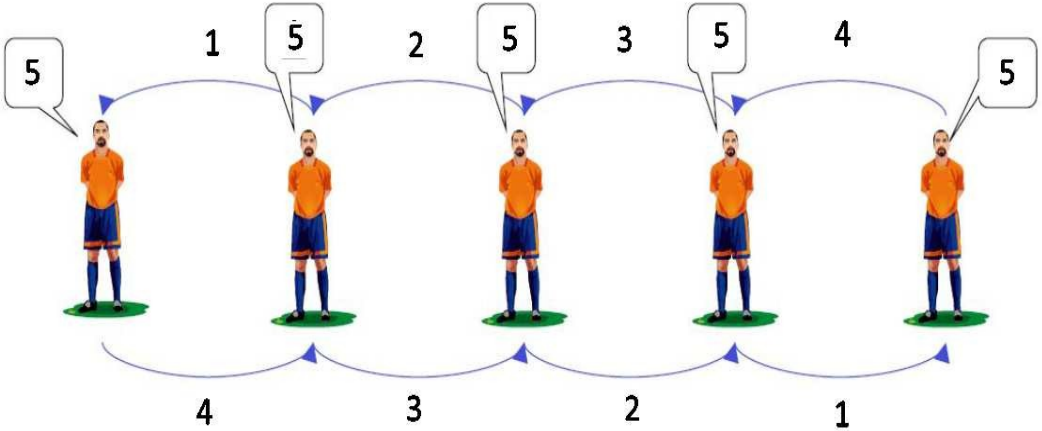


How many people are we?

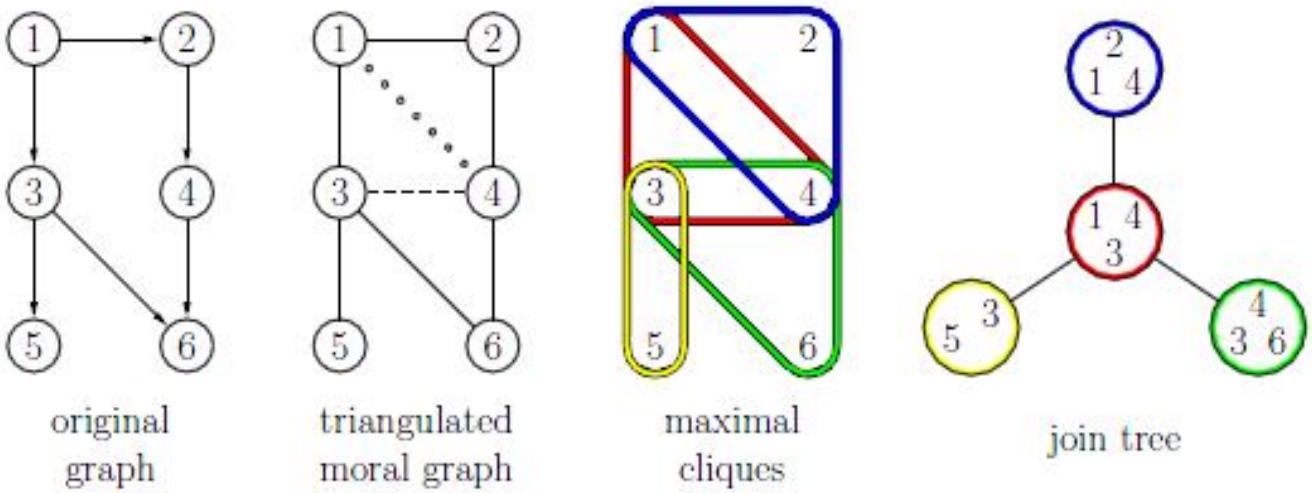
Forward propagation



backward propagation



Problem: How to avoid cycles in message pathing (as in Mycin)
Solution: Decomposition of the Probability space



- A singly connected structure is obtained by triangulating the graph and then forming a tree of maximal cliques, the so-called **join tree**.
- For evidence propagation a join tree is enhanced by so-called **separators** on the edges, which are intersection of the connected nodes → **junction tree**.

Consistent Decomposition

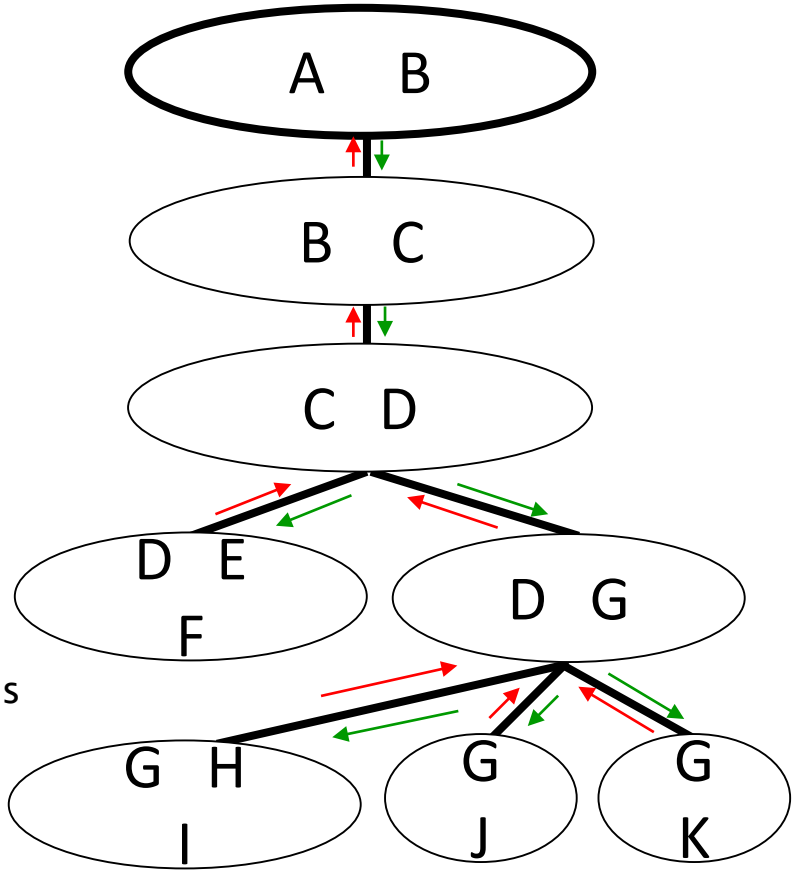
- Distributions only over the cliques (conditional independencies)
- Validity of all combinations registered and available

Fast Propagation in Join Tree

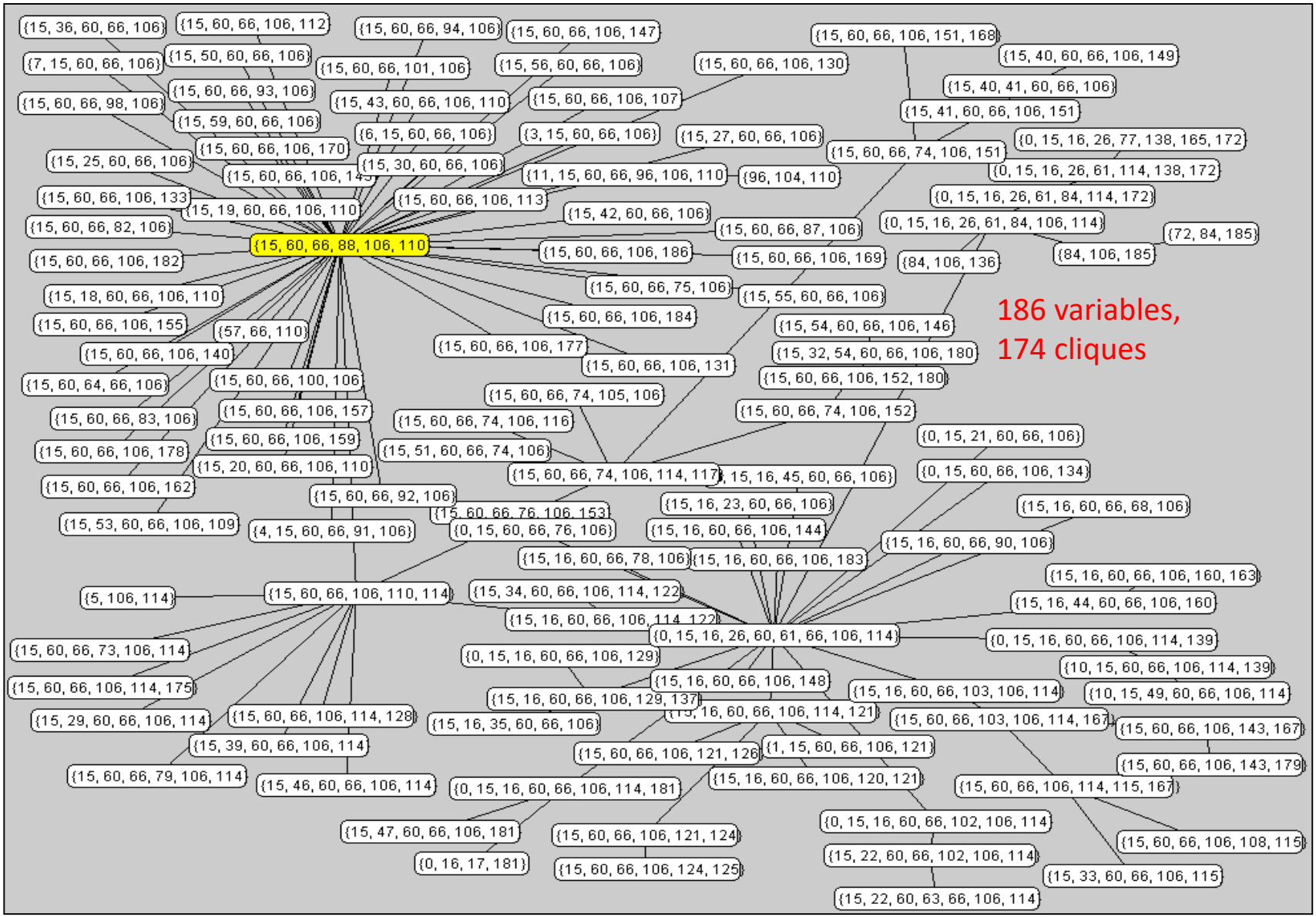
- Propagation by traversing the net twice (collect & distribute)

Efficient planning

- Planning requirements as conditional distributions
- Calculation of frequencies for arbitrary property combinations (focusing analyses, simulations) in real time



Example: Network for VW Bora



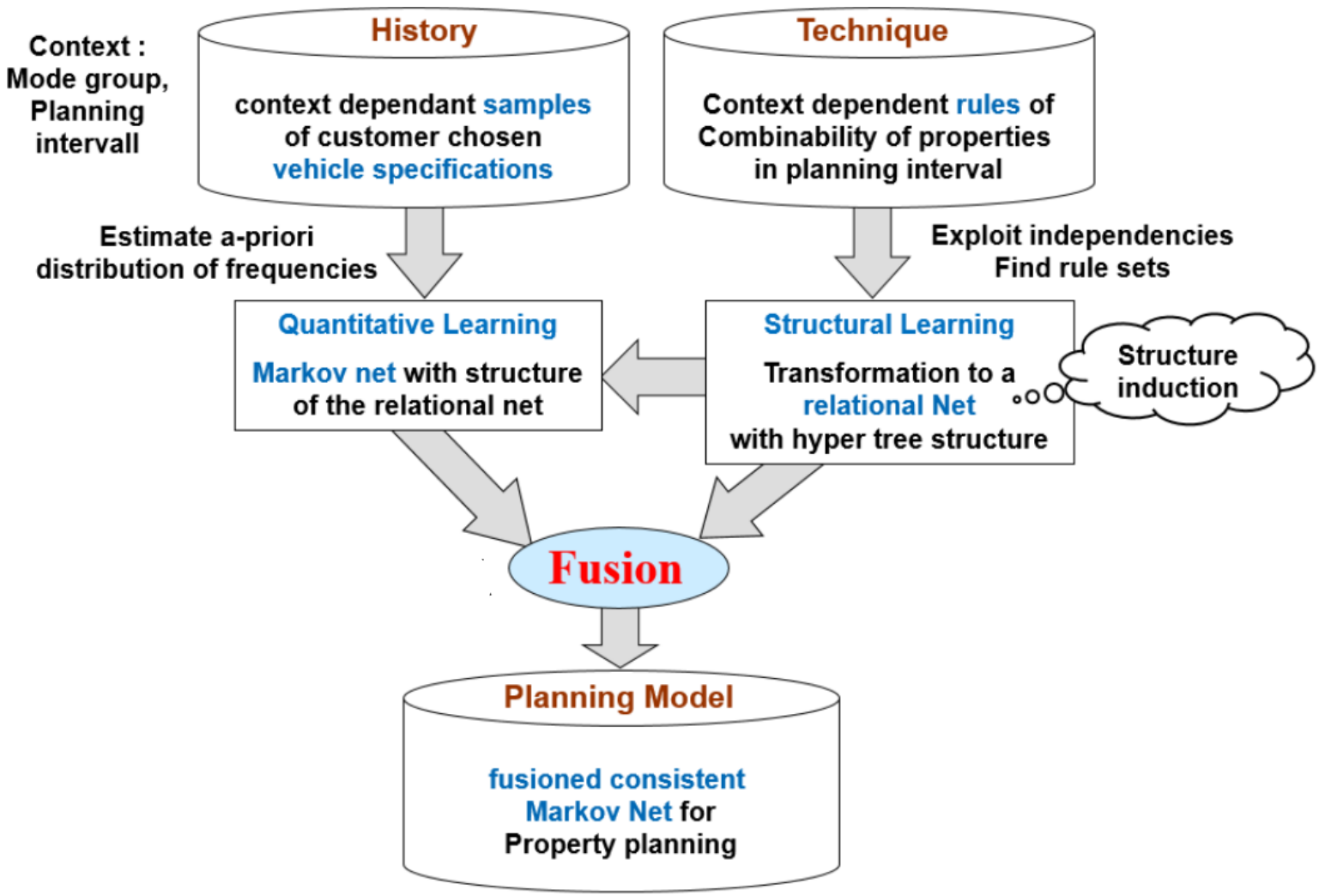
- Assistant System for Handling Estimates for Installation Rates
- Answers to the Questions in real time (seconds)
- Different Model Groups and Different Planning Intervals
- This daily use: 5000 networks (planning szenarios) handled by 350 planners worldwide
- Realization by a CI Group spin off , Leader Jörg Gebhardt

Example for one planning week

Data volume (cumulative per week)	Across all 166 Model groups	Single Model group
Number of Markov Nets	75.787	1.054
Total size (compressed)	79,9 GB	3,8 GB

… for one planning interval	Model group 1	Model group 2
Planning requirement	4.424	1.299
Number of variables	203	204
Number of cliques	174	156
Network size (compressed, RAM: ~Factor 10)	8,5 MB	17,9 MB
Largest clique (only positive probabilities)	130.806 tupels 9 variables	1.489.515 tupels 14 variables

How to learn new Planning Models?



How to revise a Planning Model?

Prior Probability Distribution

New dimensional Conditional Probabilities: Marketing Stipulations, Capacity Restrictions

REVISON

Principle of minimal Change

Posterior Probability Distribution including *specified* **and** *inferred* changes.

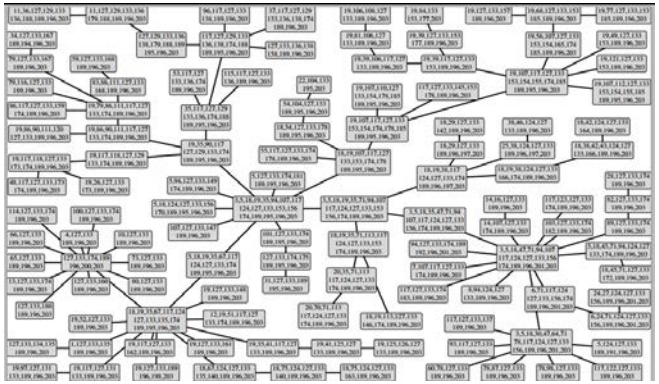
Information-theoretically closest to the prior distribution

Revision Operator

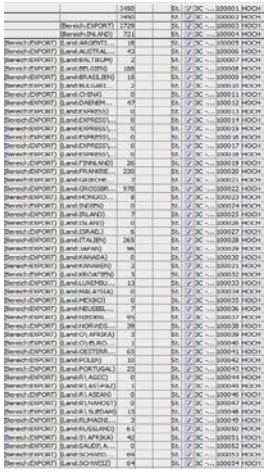
- Iterative Proportional Fitting (Biproportional Fitting, RAS Algorithm, Matrix Scaling)
- Algorithms for Adapting the Marginal Distributions
- Stepwise Modification of the Probability Distribution
- Process converge for Non-Contradicting Revision Statements
- Approx. 7000 Revision Assignments per Week and Model

Inconsistencies

Complex Structure



Many Changes



Inconsistencies

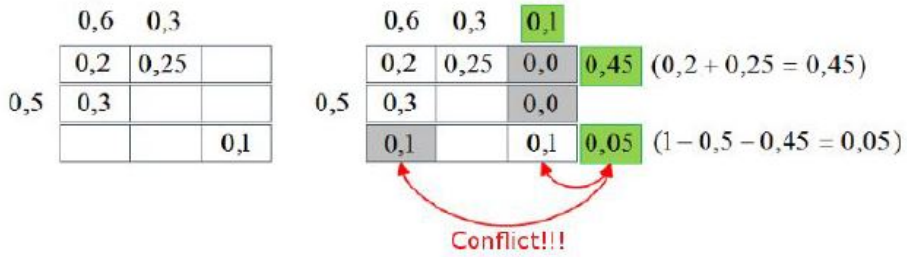
Outer Inconsistencies

Revision assigments inconsistent with zero-values in prior distribution

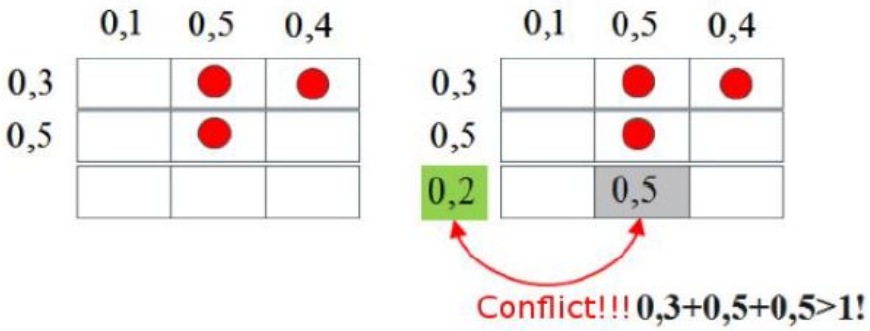
Inner Inconsistencies

Revision assignments are inconsistent, independent of prior distribution

Inconsistencies: Examples



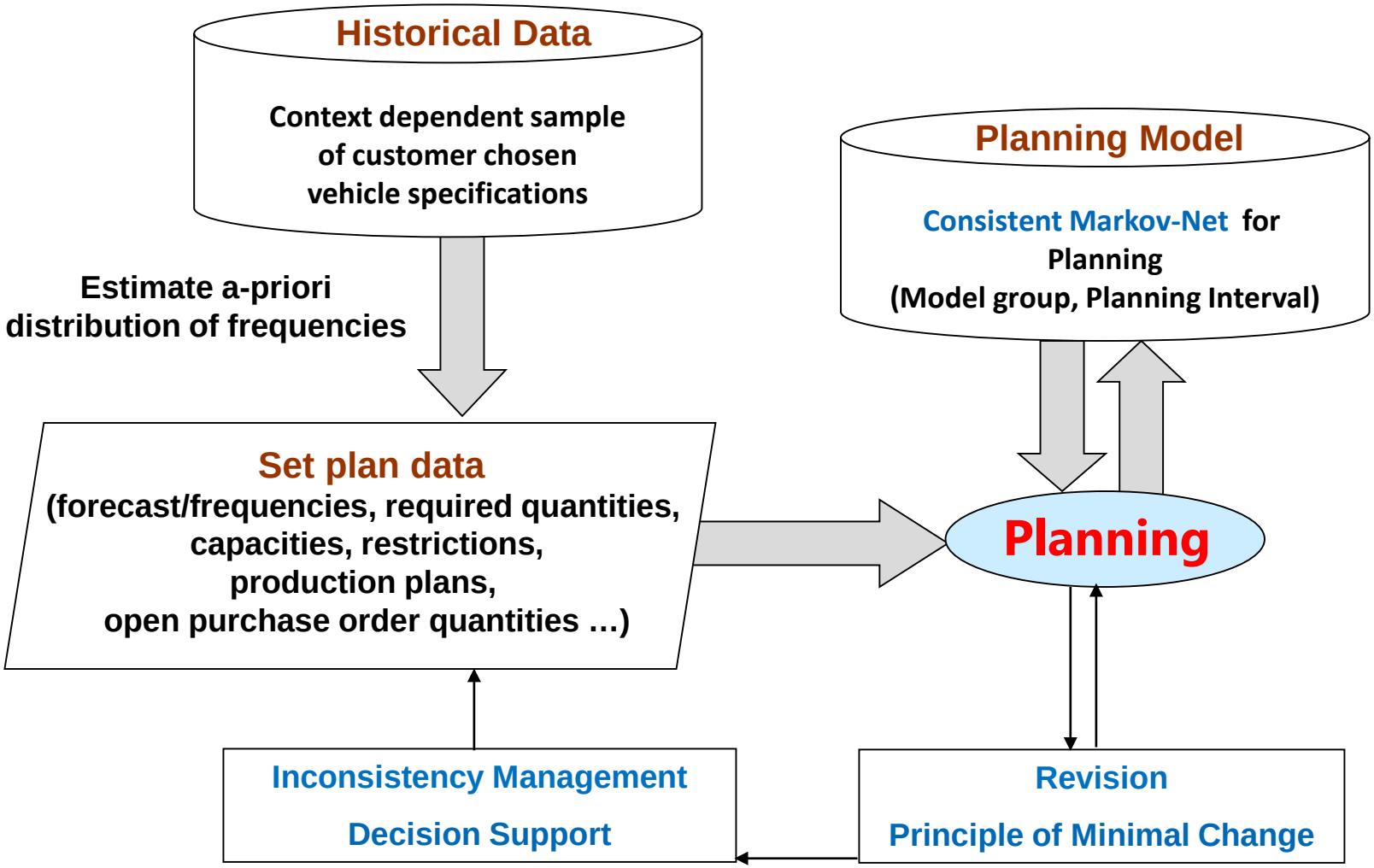
Inner Inconsistency



Outer Inconsistency

It is not easy to configure revision statements without creating inconsistencies!
How to explain the user how to change his desired revision statements?

Planning Operation: Revision



What comes next?

Associations ("observing") **Bayes, belief, conditional probability,…**
What does a symptom say about a disease? How would observing X change my belief in Y?

Intervention ("doing") **Do-Operator, model revision,…**
If I take aspirin, will my headache be cured? What would Y be if I did X?

Causality („understanding") **Counterfactuals,…**
Was it X that caused Y? What if X hadn't occurred? What if I had acted differently? Did Aspirin Stop My Headache? What if I hadn't smoked in the last 2 years?

Reference

Judea Pearl and Dana Mackenzie, The Book of Why: The New Science of Cause and Effect , 2018