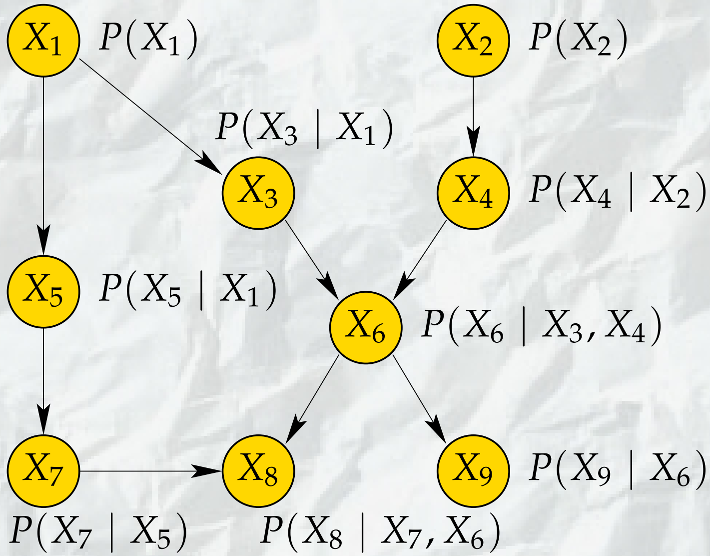
Bayesian networks in Mastermind

Jiří Vomlel

http://www.utia.cas.cz/vomlel/

Contents

- Bayesian networks, tasks solved by them, and the junction tree propagation.
- An application of BNs - adaptive testing
- The game of Mastermind as an example of adaptive test
- Bayesian networks in the game of Mastermind
- Efficient inference using Bayesian networks for Mastermind

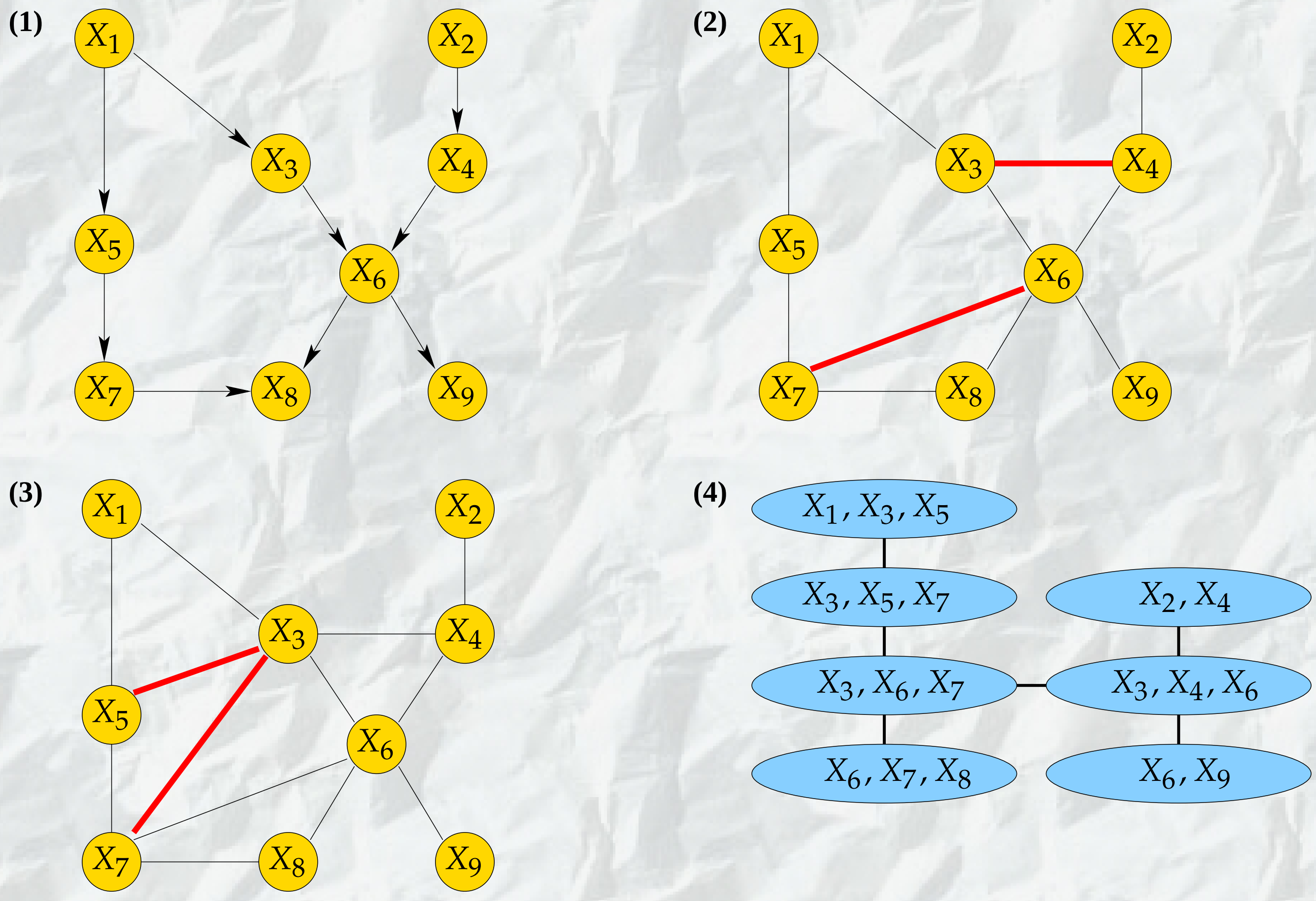


$$
\begin{aligned}
P(X_1, \ldots, X_9) \; &= \\
&= \; P(X_9|X_8, \ldots, X_1) \cdot P(X_8|X_7, \ldots, X_1) \cdot \ldots \cdot P(X_2|X_1) \cdot P(X_1) \\
&= \; P(X_9|X_6) \cdot P(X_8|X_7, X_6) \cdot P(X_7|X_5) \cdot P(X_6|X_4, X_3) \\
&\quad \cdot P(X_5|X_1) \cdot P(X_4|X_2) \cdot P(X_3|X_1) \cdot P(X_2) \cdot P(X_1)
\end{aligned}
$$

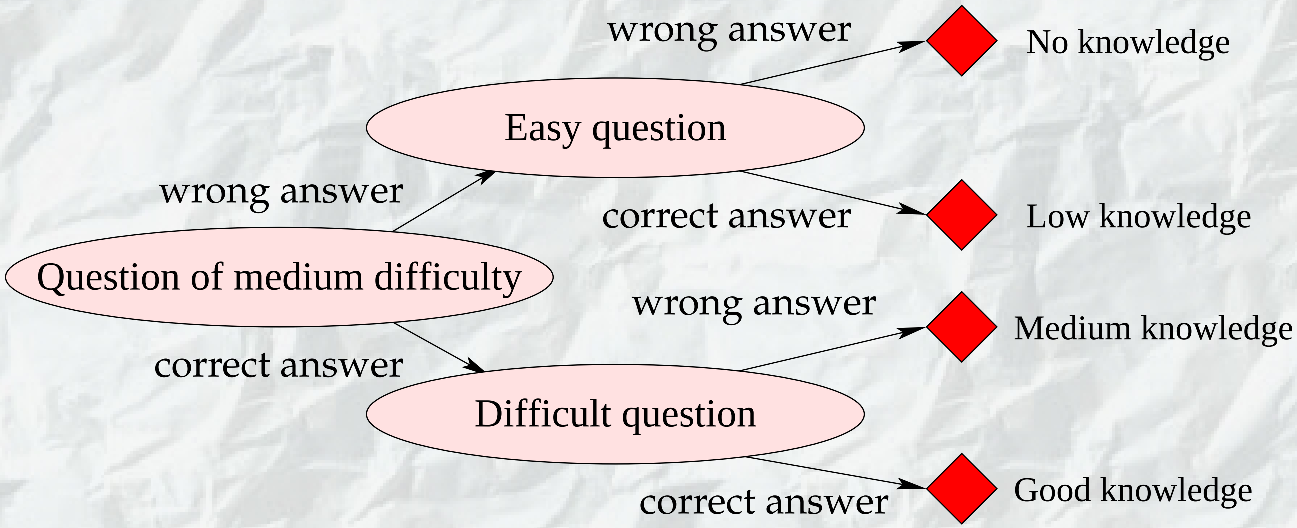
Typical use of Bayesian networks

- to model and explain a domain.
- to update beliefs about states of certain variables when some other variables were observed, i.e., computing conditional probability distributions, e.g., $P(X_{23}|X_{17} = yes, X_{54} = no)$.
- to find most probable configurations of variables
- to support decision making under uncertainty
- to find good strategies for solving tasks in a domain with uncertainty.

Bayesian networks: junction tree propagation



A simple example of an adaptive test



The game of Mastermind

T_j, H_j ... colors on the j^{th} position in the guess and in the hidden code. Let $\delta(A, B)$ equals one if $A = B$ and zero otherwise.

$$P_j = \delta(T_j, H_j) \qquad C_i = \sum_{j=1}^{4} \delta(H_j, i) \qquad M_i = \min(C_i, G_i)$$

$$P = \sum_{j=1}^{4} P_j \qquad G_i = \sum_{j=1}^{4} \delta(T_j, i) \qquad C = \left(\sum_{i=1}^{6} M_i\right) - P$$

Probability over the codes

$Q(H_1, \ldots, H_4)$ ... the probability distribution over the possible codes.
At the beginning of the game this distribution is uniform, i.e.

$$Q(H_1 = h_1, \ldots, H_4 = h_4) \quad = \quad \frac{1}{6^4} \; = \; \frac{1}{1296}$$

During the game we update probability $Q(H_1, \ldots, H_4)$ using the obtained evidence $\mathbf{e}$ and compute the conditional probability

$$Q(H_1 = h_1, \ldots, H_4 = h_4 \mid \mathbf{e}) \quad = \quad \begin{cases} \frac{1}{n(\mathbf{e})} & \text{if } (h_1, \ldots, h_4) \text{ is a possible code} \\ 0 & \text{otherwise,} \end{cases}$$

where $n(\mathbf{e})$ is the total number of codes that are possible candidates for the hidden code.

A measure of uncertainty - the Shannon entropy

A criteria suitable to measure the uncertainty about the hidden code is the Shannon entropy

$$H(Q(H_1, \ldots, H_4 \mid \mathbf{e})) \quad = \quad \sum_{h_1,\ldots,h_4} \begin{array}{c} Q(H_1 = h_1, \ldots, H_4 = h_4 \mid \mathbf{e}) \\ \cdot \log Q(H_1 = h_1, \ldots, H_4 = h_4 \mid \mathbf{e}) \end{array} \quad ,$$

where $0 \cdot \log 0$ is defined to be zero.

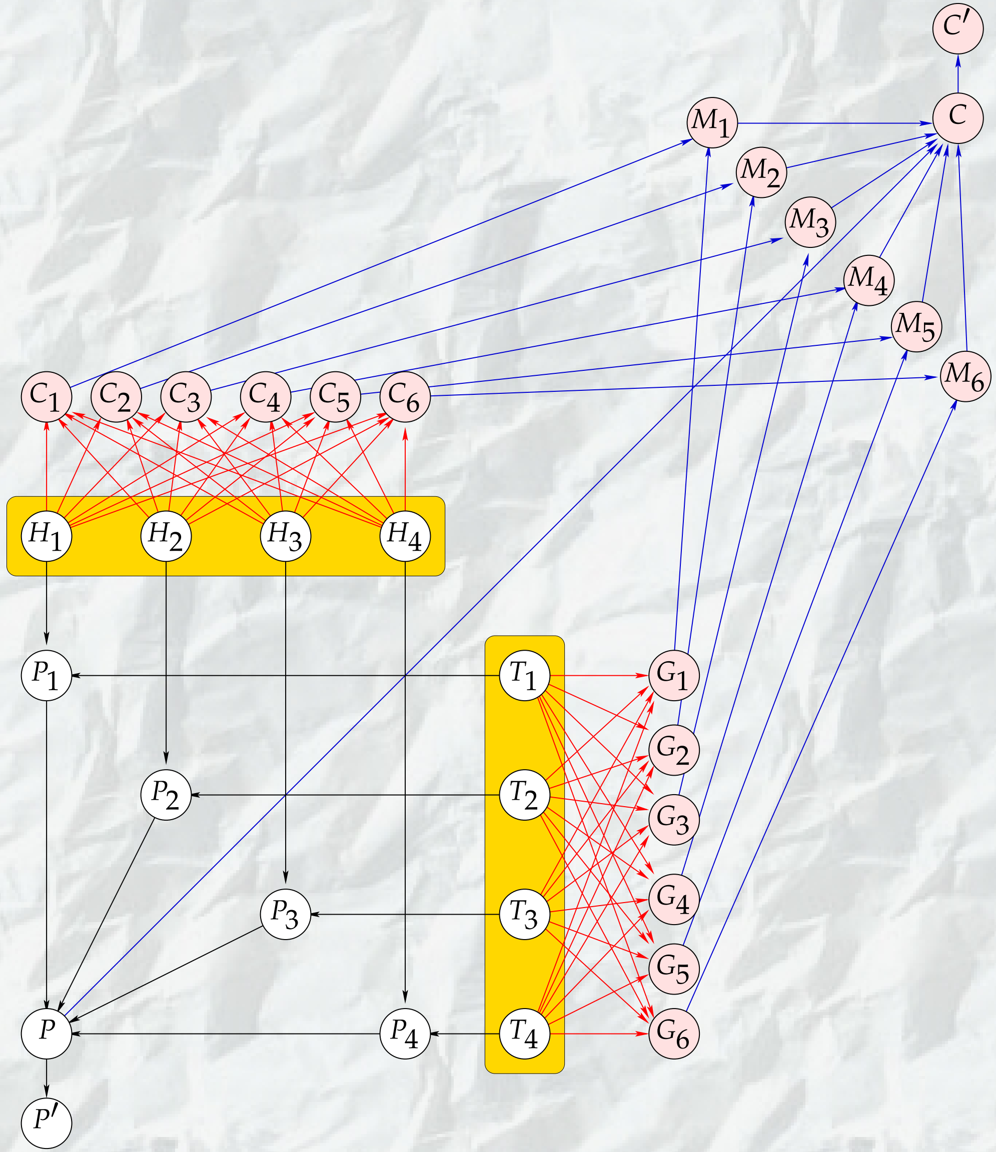
Note that the Shannon entropy is zero if and only if the code is known.

Optimal Mastermind strategies

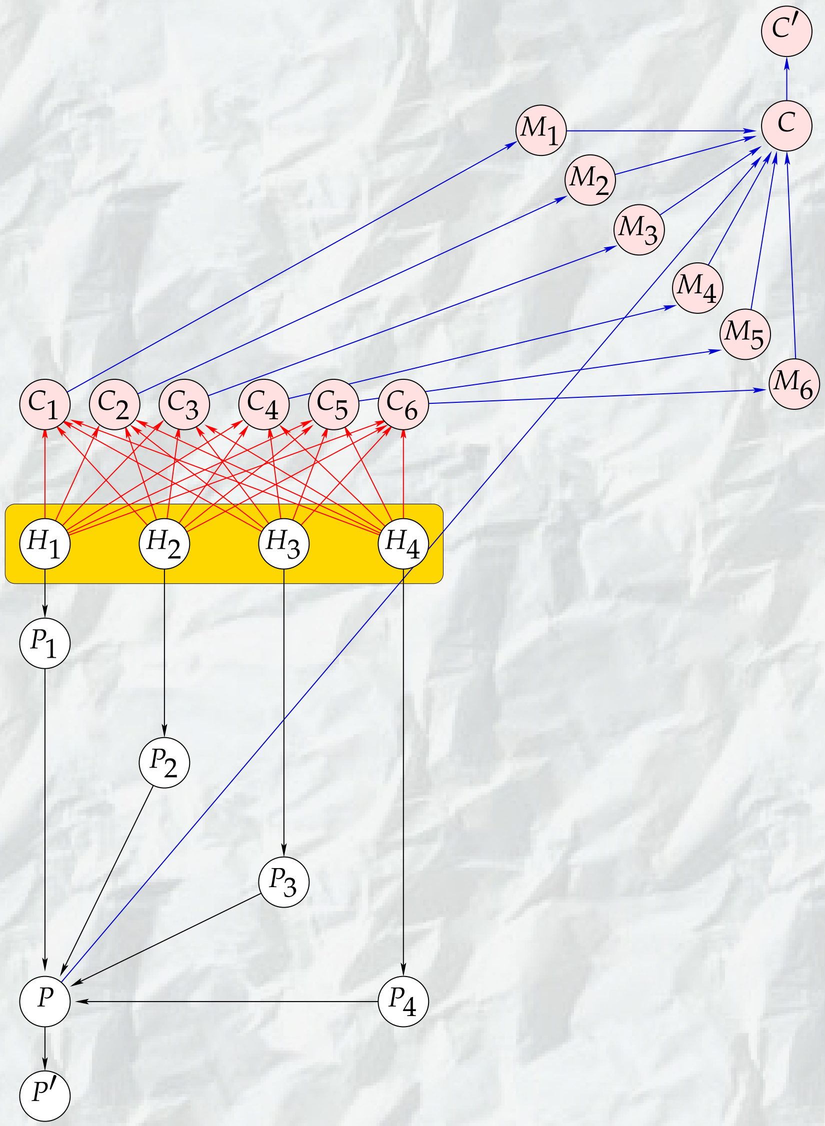
Different criteria:

- **minimal expected length**
  minimal sum over all suggested sequences of
  length of a sequence $\times$ probability of this sequence
  Koyama, Lai (1993):
  A minimal strategy with 5625/1296 = 4.340 guesses.
- **minimal depth**
  minimal number of guesses in the worst case
  Koyama, Lai (1993):
  A different strategy with depth of 5 guesses.
- **most informative within a limited number of guesses**
  minimal sum over all suggested sequences of
  entropy after a sequence $\times$ probability of this sequence

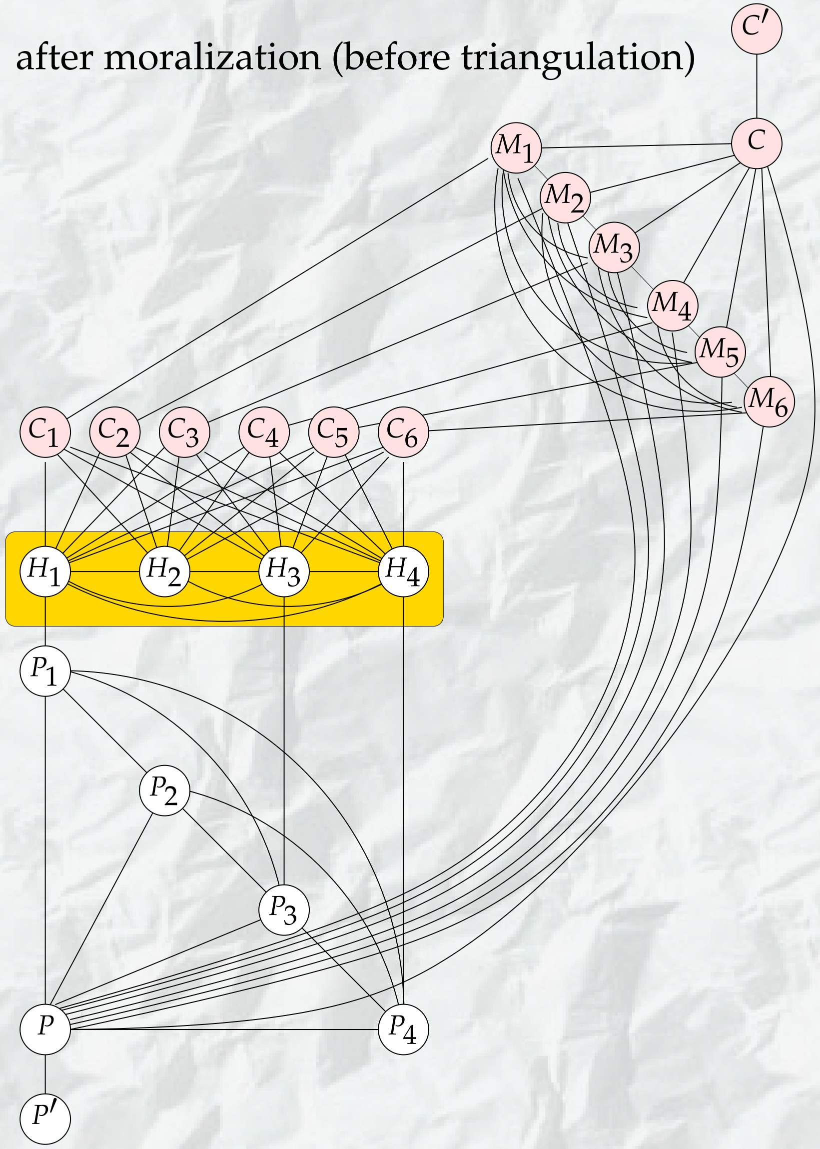
Bayesian network for the probabilistic Mastermind



Bayesian network after inserting evidence

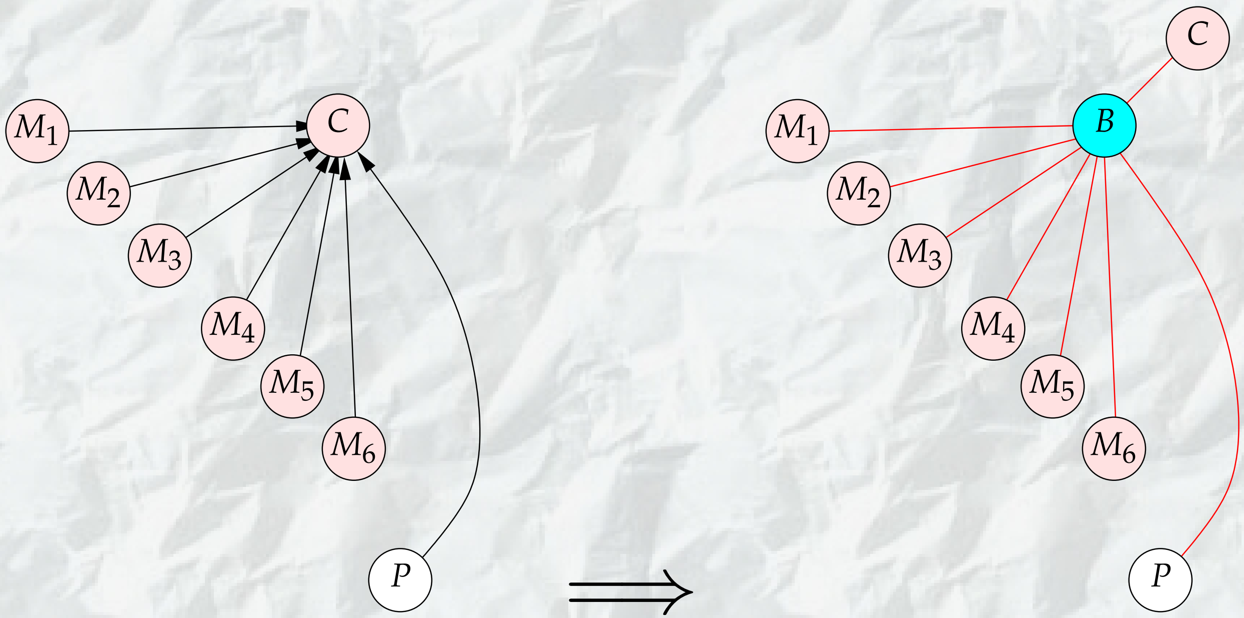


after moralization (before triangulation)

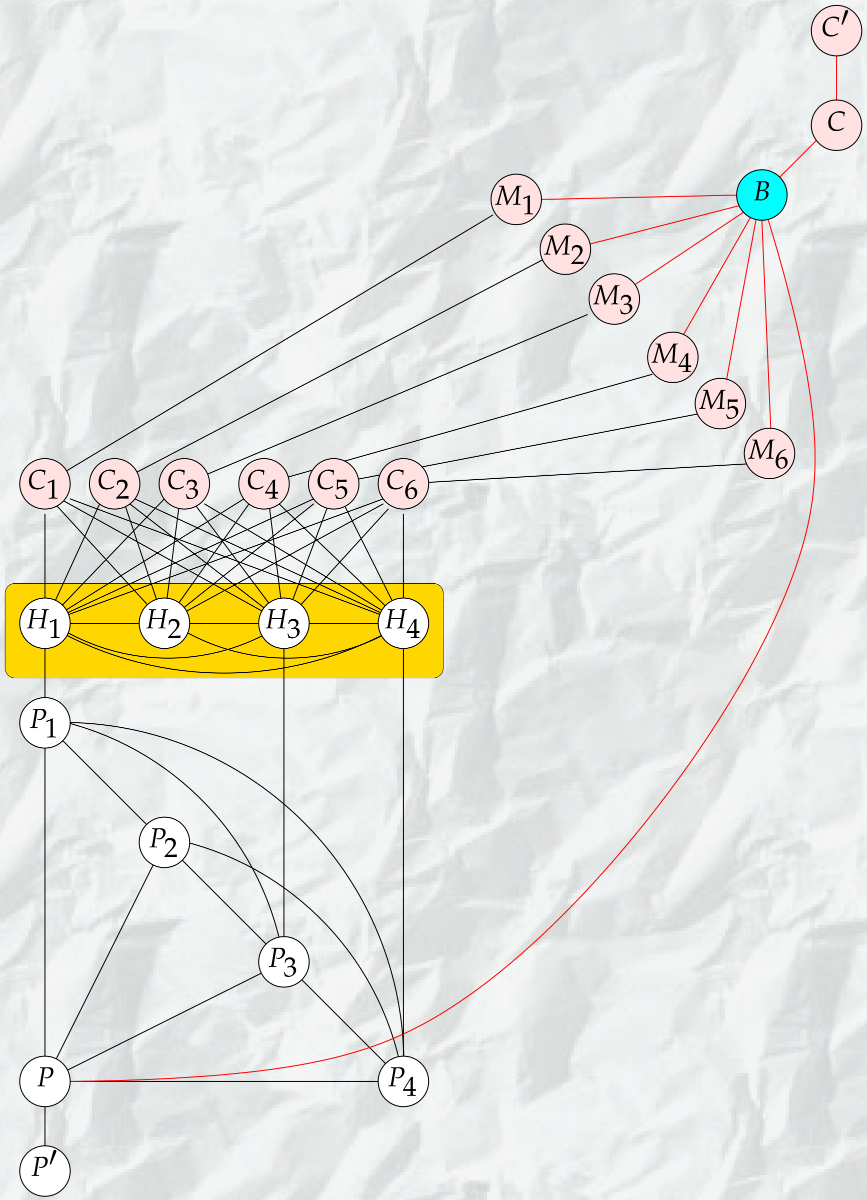


Transformation by introducing an auxiliary variable to the model

Savicky, Vomlel (2004)



Bayesian network after the suggested transformation



Junction tree size:

- without the suggested transformation $> \mathbf{20{,}526{,}445}$
- after the suggested transformation **214,775**

Summary

- The game of Mastermind is an example of a adpative test.
- In order to use Bayesian networks for computations we need to exploit functional dependences in the model.
- The suggested transformation substantially decreses computational demands.