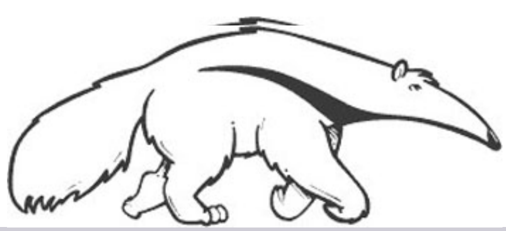
Tree decompositions & exact inference

Learning in Graphical Models

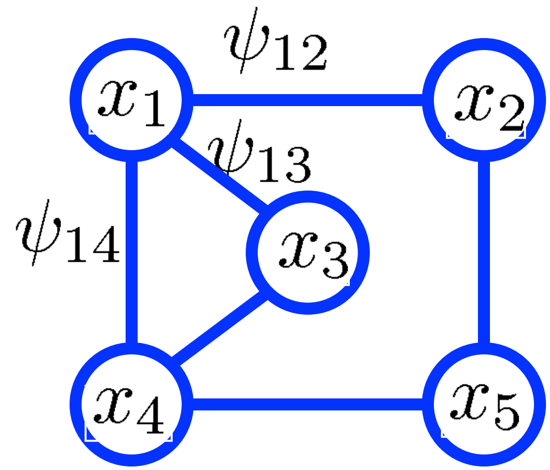
Prof. Alexander Ihler



Tree decompositions

$$p(\boldsymbol{x}) = \frac{1}{Z}\psi_{12}(x_1, x_2)\psi_{13}(x_1, x_3)\cdots$$

$$\sum_{x_2 \ldots x_5} \psi_{25}\psi_{34}\psi_{45} \sum_{x_1} \psi_{12}\psi_{13}\psi_{14}$$



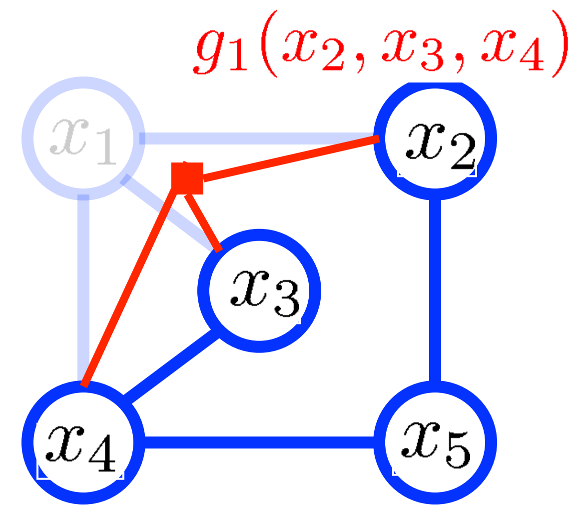
$$\{x_1, x_2, x_3, x_4\}$$

$$\begin{aligned}
B_1 &: \{\psi_{12},\ \psi_{13},\ \psi_{14}\ \} \\
B_2 &: \{\psi_{25}\ \} \\
B_3 &: \{\psi_{34}\ \} \\
B_4 &: \{\psi_{45}\ \} \\
B_5 &: \{\ \ \}
\end{aligned}$$

Tree decompositions

$$\sum_{x_3 \ldots x_5} \psi_{34}\psi_{45} \sum_{x_2} \psi_{25} \; g_1(x_2, x_3, x_4)$$

$$\begin{tikzcd} \{x_1, x_2, x_3, x_4\} \arrow[d, "g_1(x_2, x_3, x_4)"] \\ \{x_2, x_3, x_4, x_5\} \end{tikzcd}$$

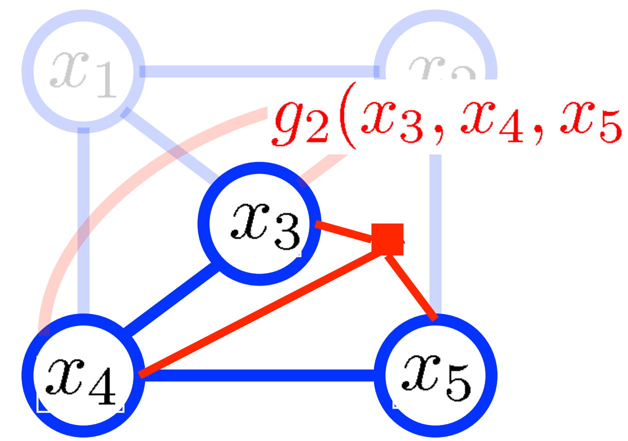


$$\begin{aligned}
B_1 &: \{\psi_{12},\ \psi_{13},\ \psi_{14}\ \} \\
B_2 &: \{\psi_{25}, g_1\ \} \\
B_3 &: \{\psi_{34}\ \} \\
B_4 &: \{\psi_{45}\ \} \\
B_5 &: \{\ \ \}
\end{aligned}$$

Tree decompositions

$$\sum_{x_3 \dots x_5} \psi_{34}\psi_{45}\ g_2(x_3, x_4, x_5)$$

$$\begin{tikzcd} \psi_{12}\psi_{13}\psi_{14} \arrow[r] & \{x_1, x_2, x_3, x_4\} \arrow[d, "g_1(x_2, x_3, x_4)"] \\ \psi_{25} \arrow[r] & \{x_2, x_3, x_4, x_5\} \arrow[d, "g_2(x_3, x_4, x_5)"] \\ \psi_{34}\psi_{45} \arrow[r] & \{x_3, x_4, x_5\} \end{tikzcd}$$



$B_1 \ : \ \{\psi_{12},\ \psi_{13},\ \psi_{14}\ \}$

$B_2 \ : \ \{\psi_{25}, g_1\ \}$

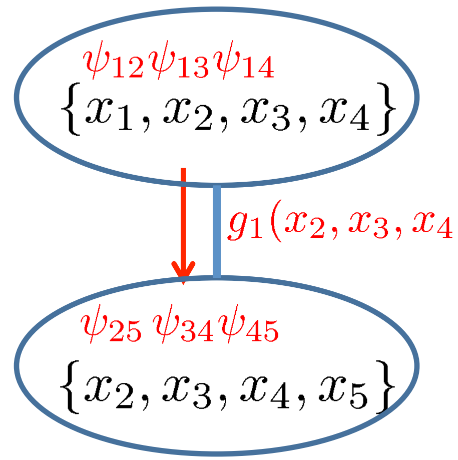
$B_3 \ : \ \{\psi_{34}, g_2\ \}$

$B_4 \ : \ \{\psi_{45}\ \}$

$B_5 \ : \ \{\ \ \}$

Tree decomposition

- Tree-structured hyper-graph
  - "Junction tree", "clique tree", "join tree" …
  - Tree over sets of nodes
  - "Separator sets"

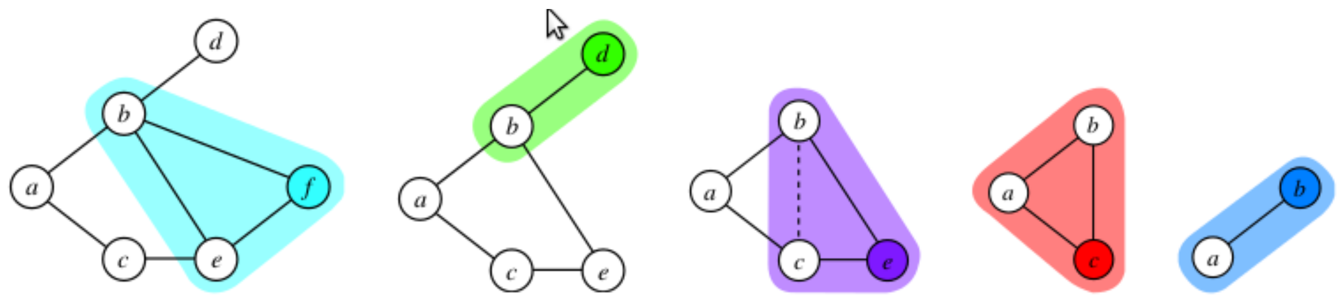


- Properties
  - Singly connected
  - Covering: every factor included in with some clique
  - Running intersection: subgraph of each variable is connected:
    Every pair containing xi has a path also containing xi

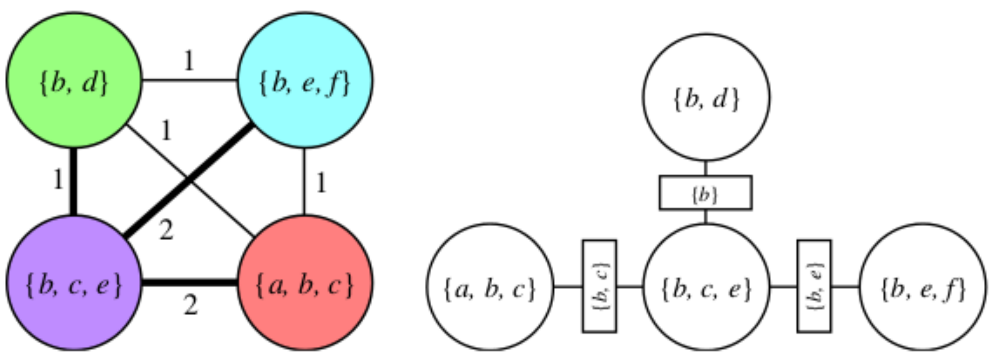
- Elimination order $\Leftrightarrow$ junction tree

Constructing a junction tree

- Choose an elimination order & triangulate the graph
  - Remember cliques created during each elimination step



- Score pairs of cliques by size of their intersection
- Find a max-weight spanning tree

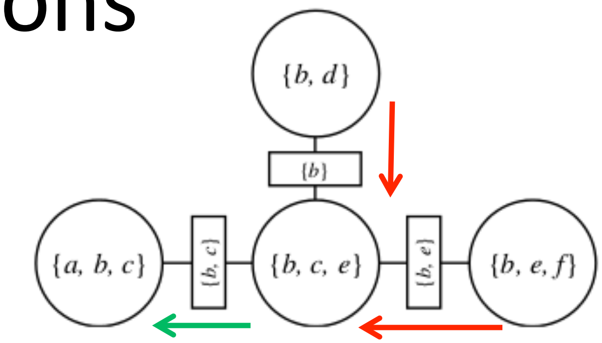


(Figures by M. Paskin)

Inference in tree decompositions

- Pass messages to root

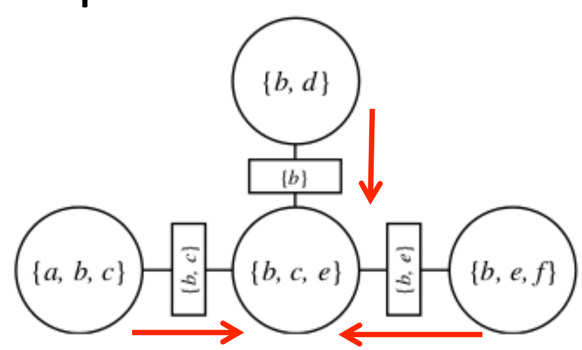
$$m_{\alpha\beta}(x_\beta) = \sum_{x_\alpha \setminus x_\beta} \psi_\alpha(x_\alpha) \prod_{\gamma \neq \beta} m_{\gamma\alpha}(x_\alpha)$$



- Usually, 2-pass procedure (leaves to root, root to leaves)
  - Computes messages needed for *any* choice of root node

- Compute normalization constant or marginal probabilities

$$p_\alpha(x_\alpha) \propto \psi_\alpha(x_\alpha) \prod_{\gamma} m_{\gamma\alpha}(x_\alpha)$$



- Computation scales with largest clique
  - Discrete: exponential; Gaussian: cubic

(Figures by M. Paskin)

Aside: decomposable models

- A model M is decomposable if there exists a tree decomposition for M with exactly the same cliques

- Equivalently,
  - M is already triangulated
  - M is already sufficient for exact inference
  - …

Learning with approximate inference

- Since inference is hard, but we need it in learning…
- Substitute some simpler algorithm?

- Monte Carlo approximations
  - Use random sampling and simulation to estimate probabilities

- Variational approximations
  - Convert inference to an optimization problem & approximate