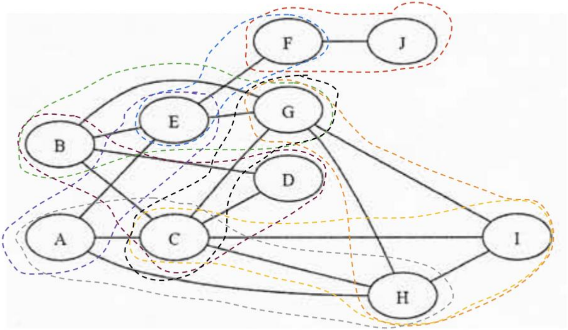
# Directed and Undirected Graphs

**Robert Schüle, Christoph Ende**

TU Berlin
Machine Intelligence I

January 26, 2016




| | | |
|---|---|---|
| $C_1$ : $F, J$ | $C_4$ : $G, H, I$ | $C_7$ : $B, C, D$ |
| $C_2$ : $F, E$ | $C_5$ : $C, H, I$ | $C_8$ : $A, E$ |
| $C_3$ : $B, E, G$ | $C_6$ : $A, C, H$ | $C_9$ : $C, G$ |

# a) construction of the moral graph

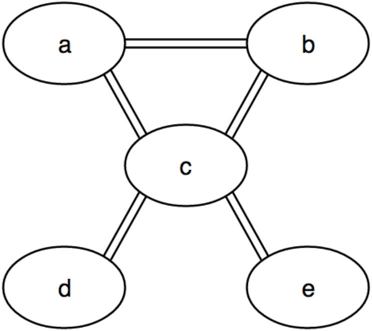



Figure: moral graph of given DAG

# b) pdf from cliques and seperators

- we show that the DAG representation equals the given formula for the joint distribution:

$$
\begin{aligned}
P(a, b, c, d, e) &= P(a)P(b)P(c|a, b)P(d|c)P(e|c) && (1)\\
&= \frac{P(a, b)P(a, b, c)P(d, c)P(e, c)}{P(c)^2 P(a, b)} && (2)\\
&= \frac{P(a, b, c)P(d, c)P(e, c)}{P(c)P(c)} && (3)\\
&= \frac{P(C_1)P(C_2)P(C_3)}{P(S_1)P(S_2)} = \frac{\prod_i P(C_i)}{\prod_j P(S_j)} && (4)\\
& && (5)
\end{aligned}
$$

# a) potential functions and the clique-marginal representation

- potential functions are used to represent joint probabilities of cliques, can be easily updated by new data
- potential functions are strictly positive, real-valued and only depend on random variables in their assigned clique
- don't need to be normalised so that $\psi_n(C_n) = \alpha\, P(C_n),\ \alpha > 0$
- link between joint probability distribution and clique-marginal representation:

$$P(V \in \mathcal{V}) \;=\; \prod_{V \in \mathcal{V}} P(V|V_{pa(V)}) \;=\; \frac{\prod_i P(C_i)}{\prod_j P(S_j)} \;=\; \alpha \frac{\prod_i \psi_i(C_i)}{\prod_j \psi_j(S_j)} \qquad (6)$$

- at the beginning, all potentials are chosen to be 1

# b) exploitation of potential functions to make inferences

1. update clique potential
   - inference by observed evidence at one node: $V_l = v_l$
   - introduce indicator function $E(V_l) = \begin{cases} 1 & \text{se } V_l = v_l \\ 0 & \text{se else} \end{cases}$
   - update clique potential of observed node: $\psi_{i,new}(C_i) = \psi_i(C_i)E(V_l)$
2. message passing
   - new seperator potential: $\psi_c^*(S_c) \ = \ \sum_{V \in C_i, V \notin \psi_c} \psi_i(V \in C_i)$
   - the update of one clique demands updates of the neighbor cliques which happens by passing the update ratio $\frac{\psi_c^*(S_c)}{\psi_c(S_c)}$ ($S_c$: common seperator)
   - updated neighbor potential obtained by multiplication of current potential with the update rate

# c) benefit in comparison to DAG factorisation

- simpler and clearer models due to substitution of subgraphs by cliques
- highly improved interference

# Literature

'Machine Intelligence I', Lecture Notes by Prof. Dr. Klaus Obermayer, 2015

'Probabilistic Networks and Expert Systems', Chapter 3, Cowell et al., 1999