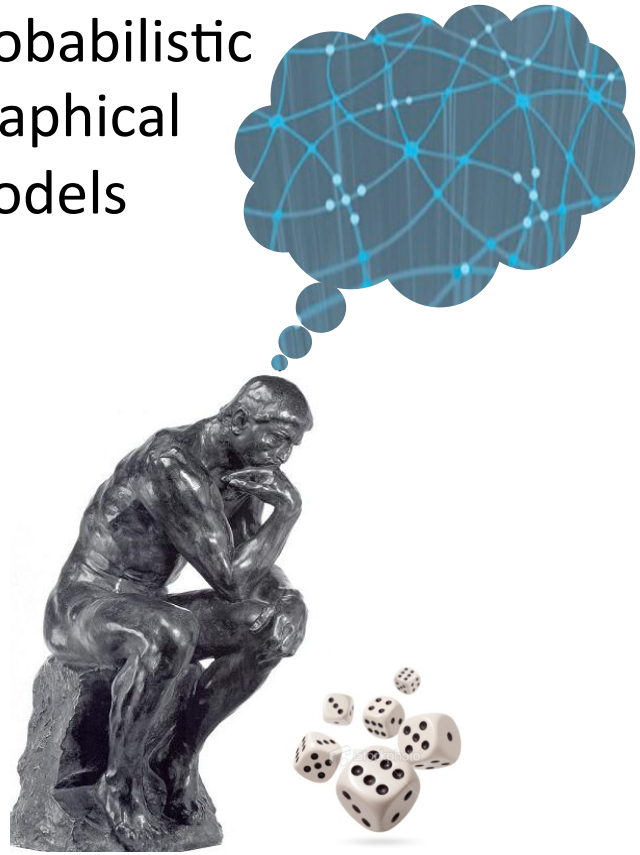
Probabilistic Graphical Models

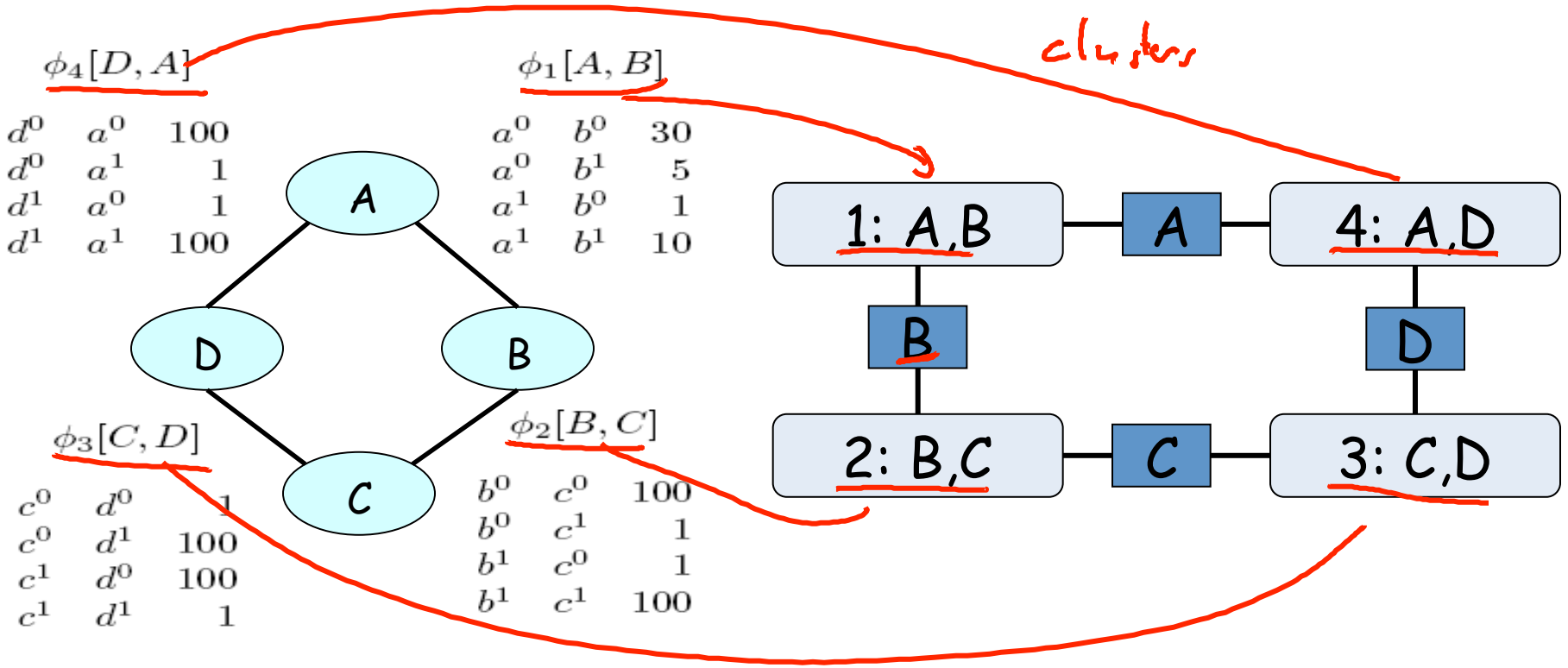


Inference

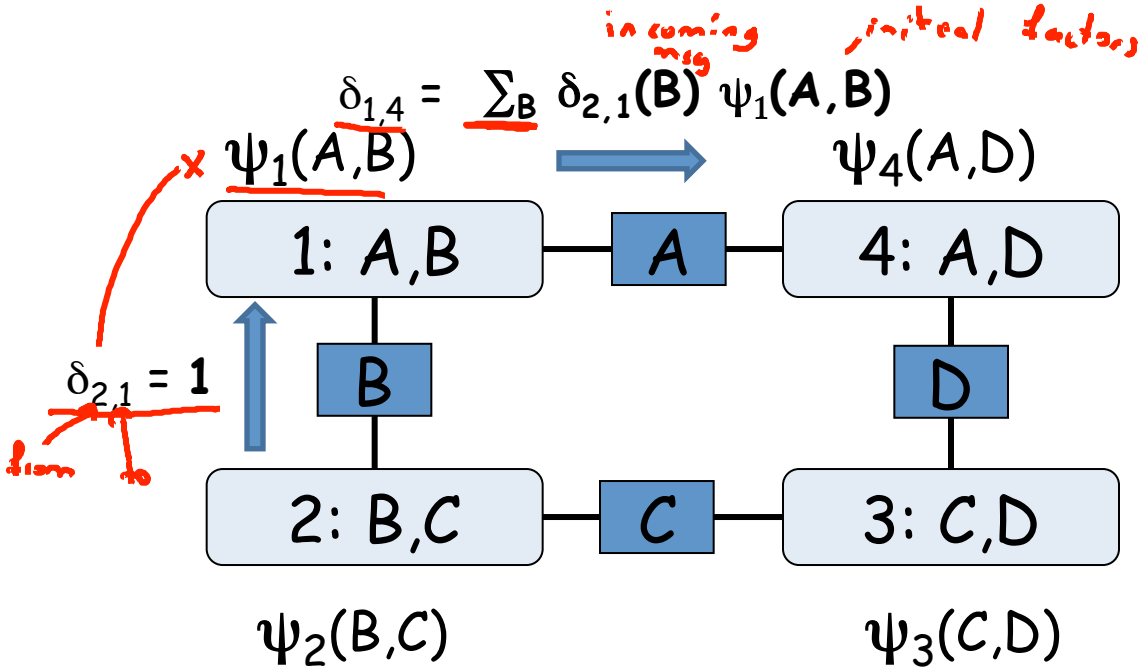
Message Passing

Belief Propagation Algorithm

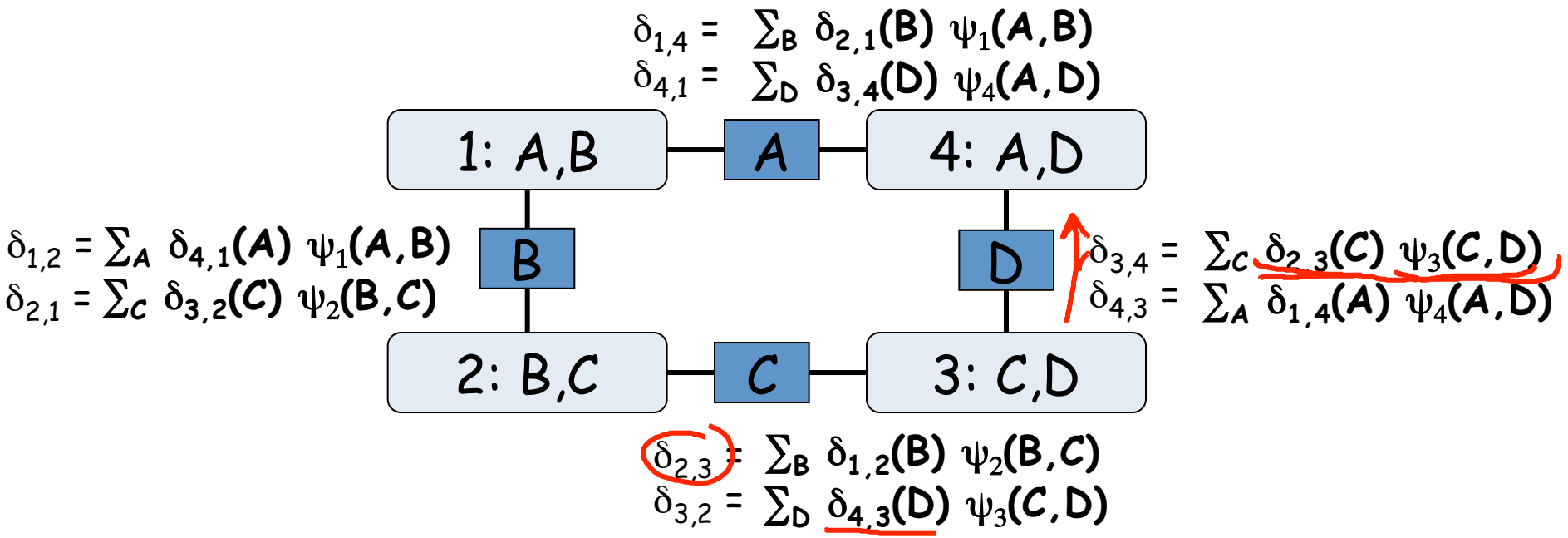
Cluster Graph



Passing Messages



Passing Messages



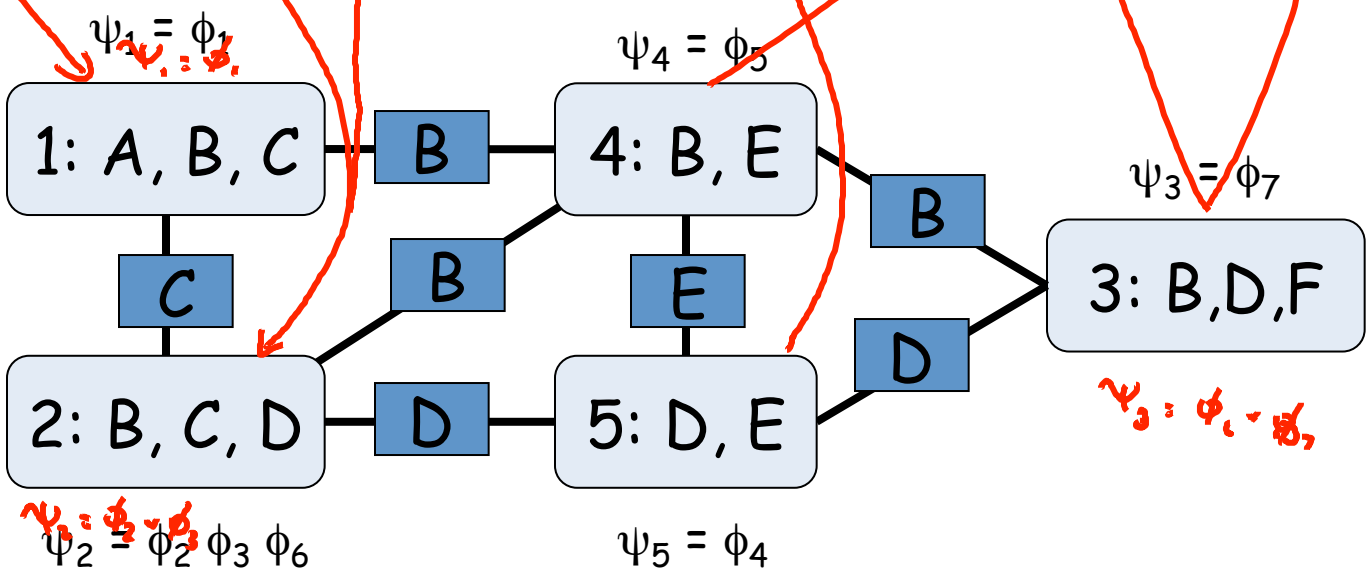
Daphne Koller

Cluster Graphs

- Undirected graph such that:
  - nodes are clusters $\boldsymbol{C}_i \subseteq \{X_1,\ldots,X_n\}$ — subsets of variables
  - edge between $\boldsymbol{C}_i$ and $\boldsymbol{C}_j$ associated with sepset $\boldsymbol{S}_{i,j} \subseteq \boldsymbol{C}_i \cap \boldsymbol{C}_j$ ← variables that they talk about
- Given set of factors Φ, we assign each ϕ_k to a cluster $\boldsymbol{C}_{\alpha(k)}$ s.t. Scope$[\phi_k] \subseteq \boldsymbol{C}_{\alpha(k)}$
- Define $$\psi_i(C_i) = \prod_{k:\alpha(k)=i} \phi_k$$ — subset; all factors assigned to it

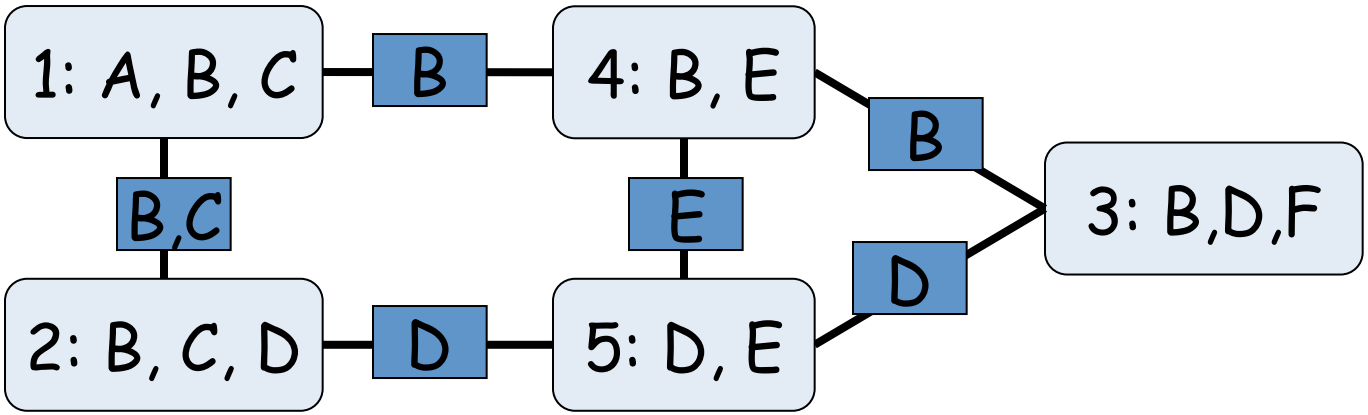
Example Cluster Graph

$\phi_1(A,B,C), \phi_2(B,C), \phi_3(B,D), \phi_4(D,E), \phi_5(B,E), \phi_6(B,D), \phi_7(B,D,F)$



Different Cluster Graph

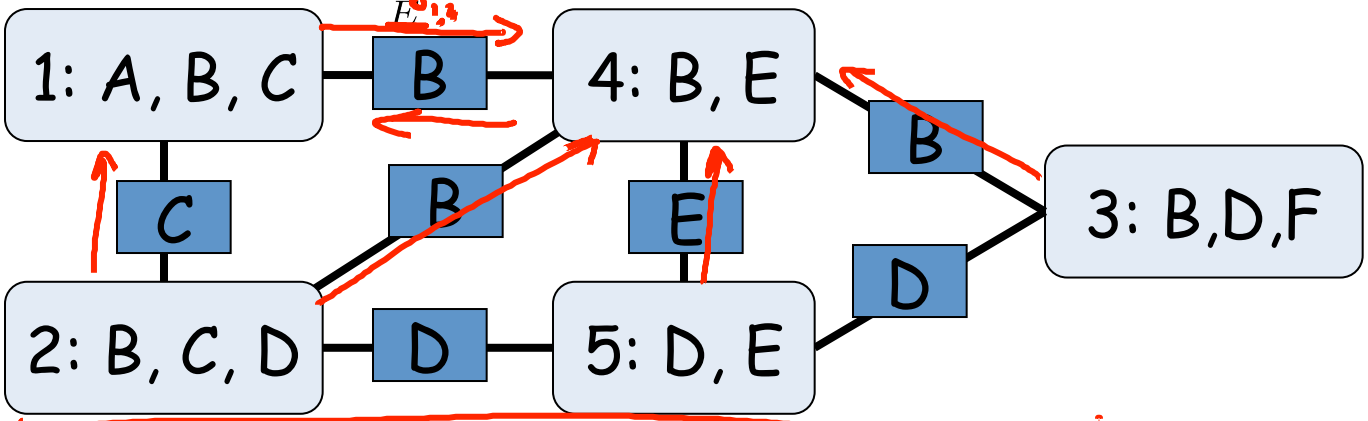
$$\phi_1(A,B,C), \phi_2(B,C), \phi_3(B,D), \phi_4(D,E), \phi_5(B,E), \phi_6(B,D), \phi_7(B,D,F)$$



Message Passing

$$\delta_{1\to 4}(B) = \sum_{A,C} \psi_1(A,B,C)\delta_{2\to 1}(C)$$

$$\delta_{4\to 1}(B) = \sum_{E} \psi_4(B,E) \times \delta_{2\to 4}(B) \times \delta_{5\to 4}(E) \times \delta_{3\to 4}(B)$$



$$\delta_{i\to j}(\boldsymbol{S}_{i,j}) = \sum_{\boldsymbol{C}_i - \boldsymbol{S}_{i,j}} \psi_i \times \prod_{k\in(\mathcal{N}_i - \{j\})} \delta_{k\to i}$$

Belief Propagation Algorithm

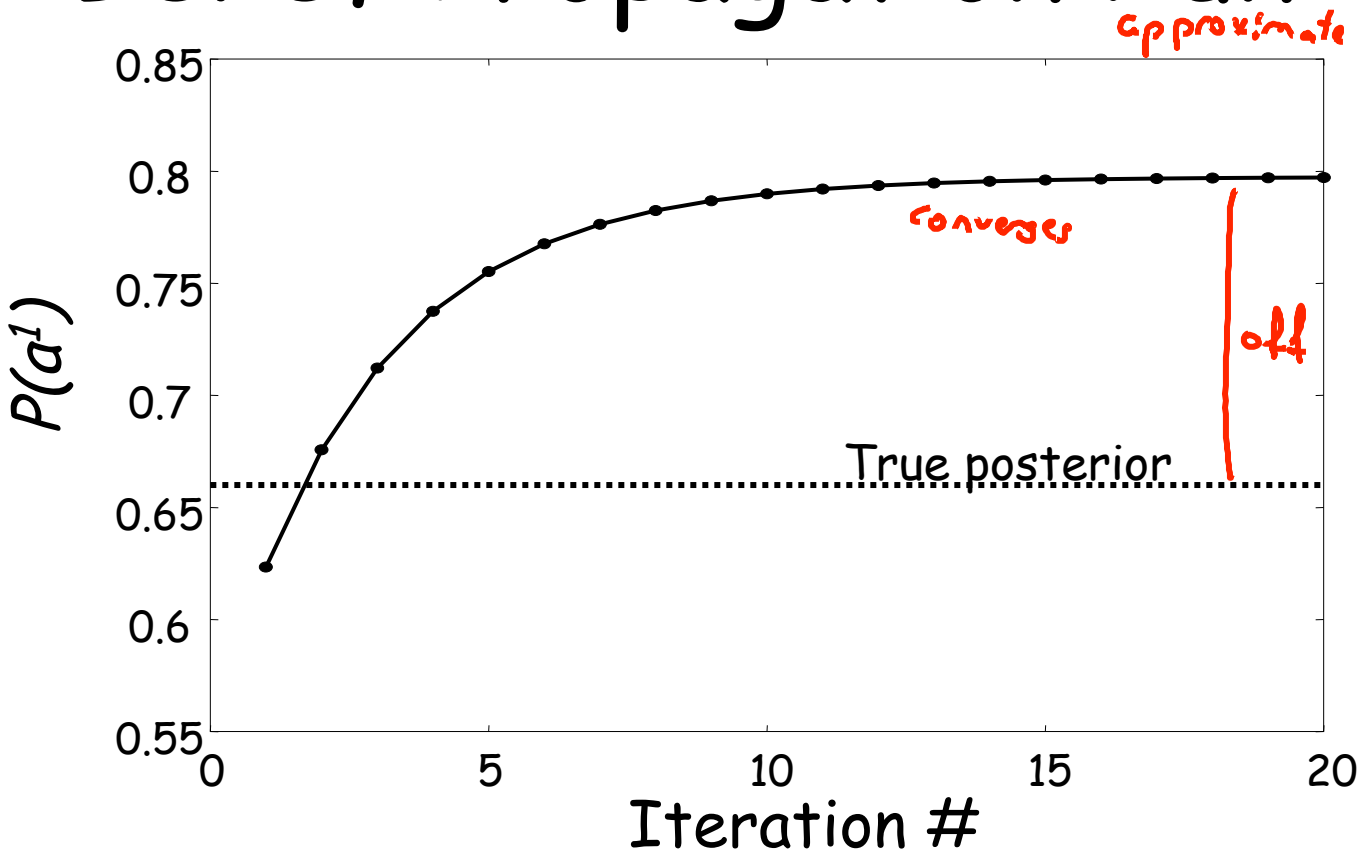
- Assign each factor $\phi_k \in \Phi$ to a cluster $C_{\alpha(k)}$
- Construct initial potentials $\psi_i(C_i) = \prod_{k:\alpha(k)=i} \phi_k$
- Initialize all messages to be **1**
- Repeat until when?
  - Select edge (i,j) and pass message round robin

$$\delta_{i \to j}(S_{i,j}) = \sum_{C_i - S_{i,j}} \psi_i \times \prod_{k \in (\mathcal{N}_i - \{j\})} \delta_{k \to i}$$

- Compute $\beta_i(C_i) = \psi_i \times \prod_{k \in \mathcal{N}_i} \delta_{k \to i}$

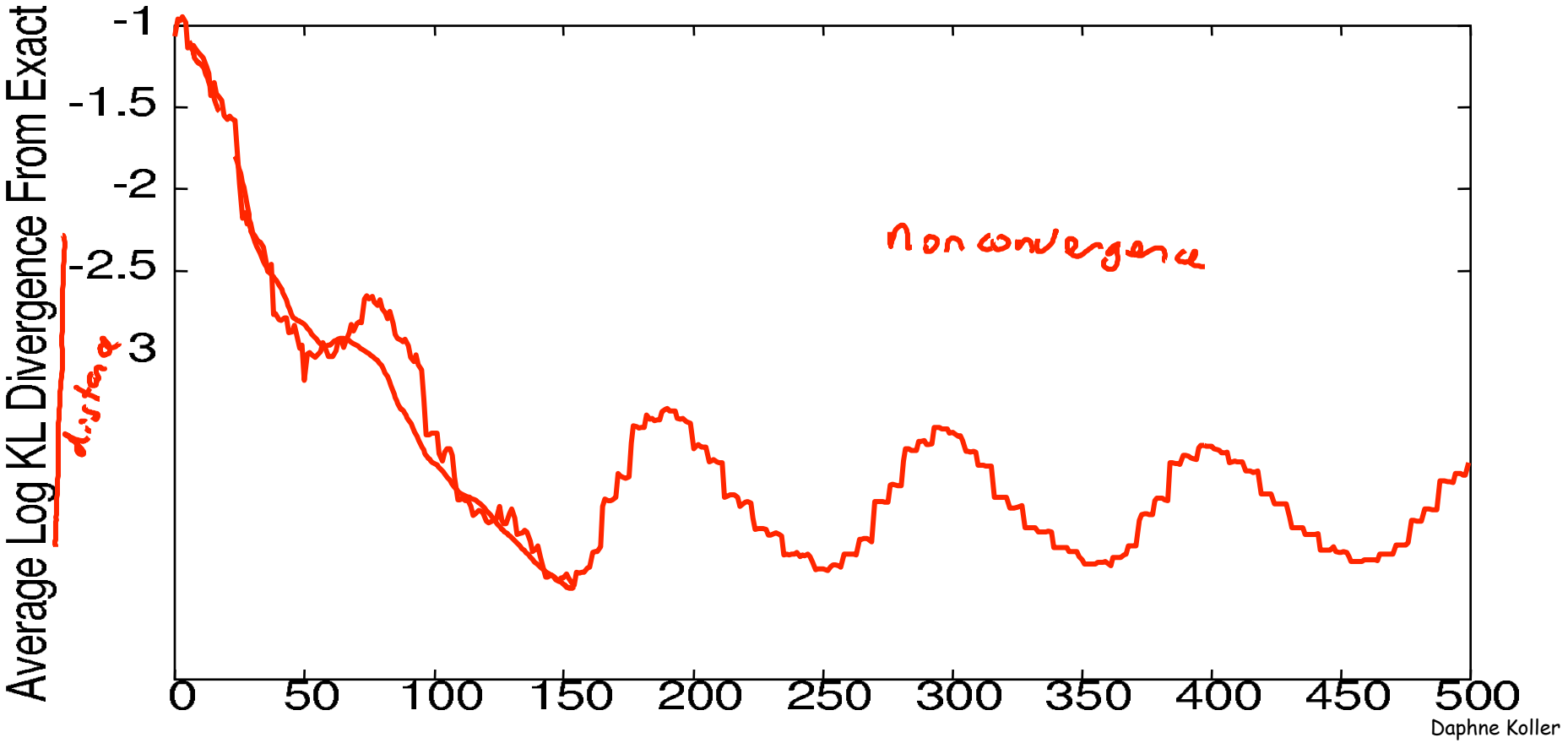
Beliefs — all neighbors

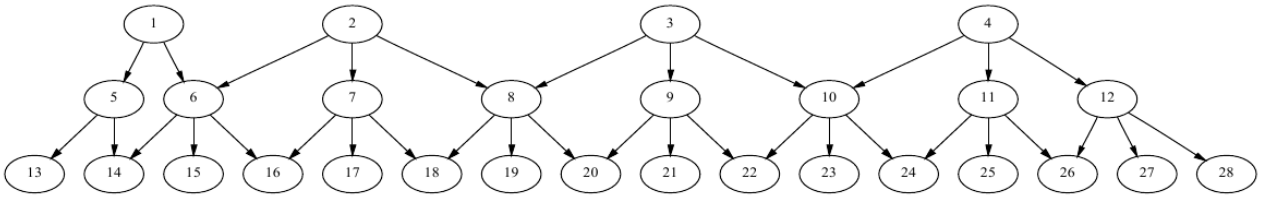
Belief Propagation Run



Daphne Koller

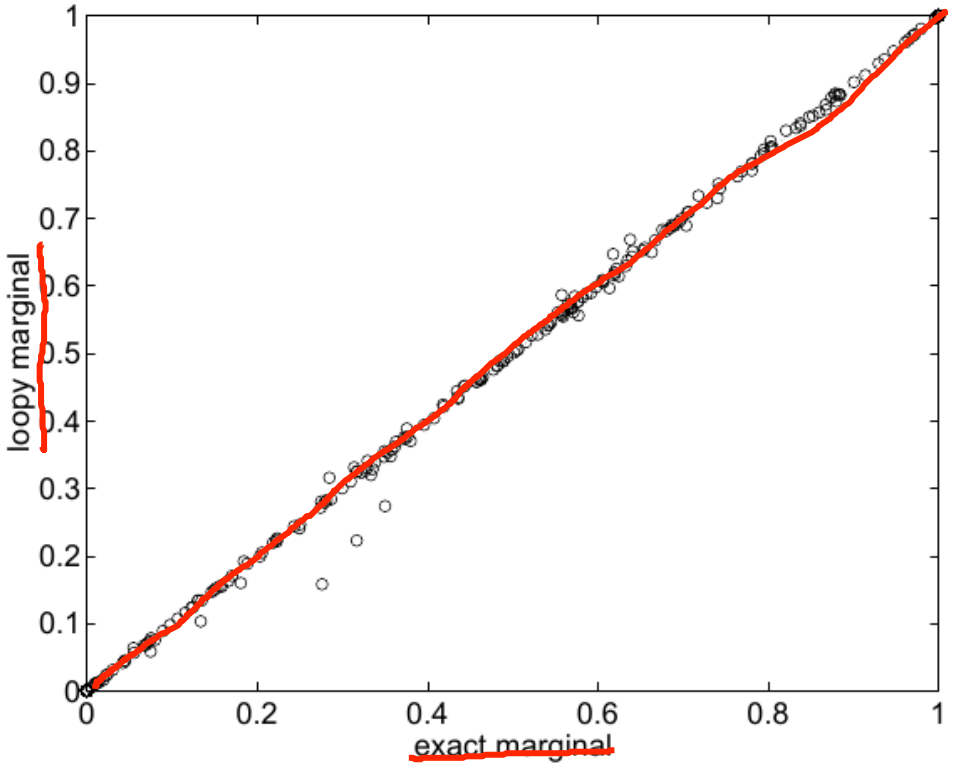
Different BP Run



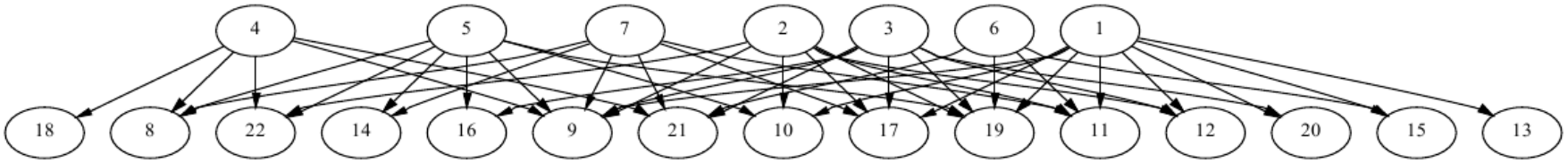


Sample Results

Murphy, Weiss, Jordan
UAI 99

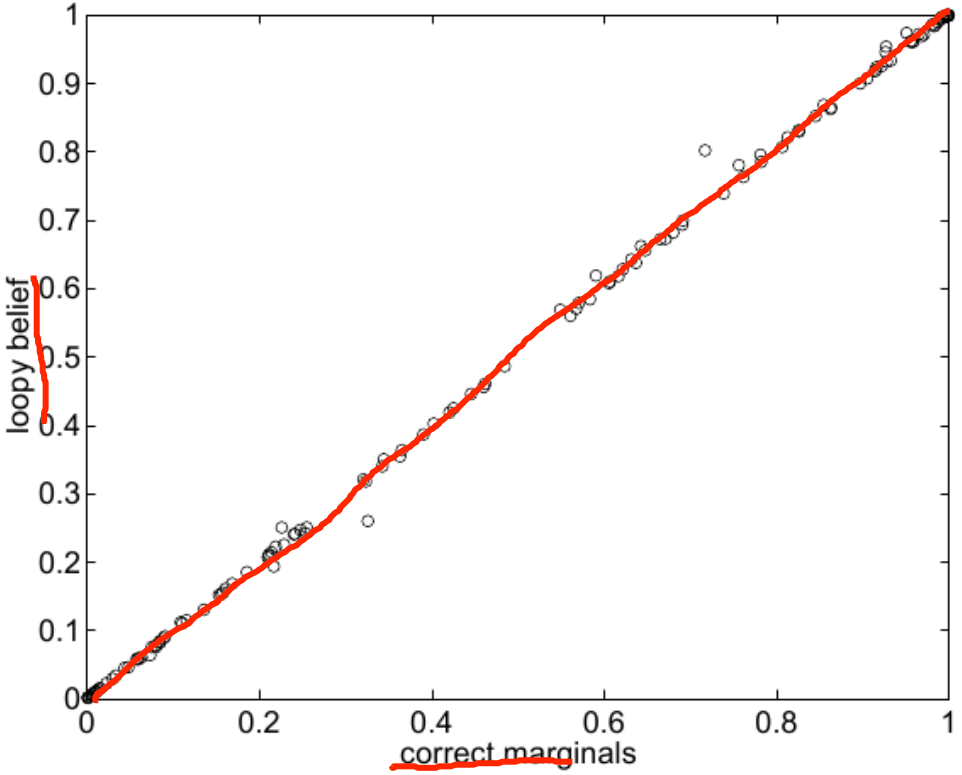


Daphne Koller



Sample Results

Murphy, Weiss, Jordan
UAI 99

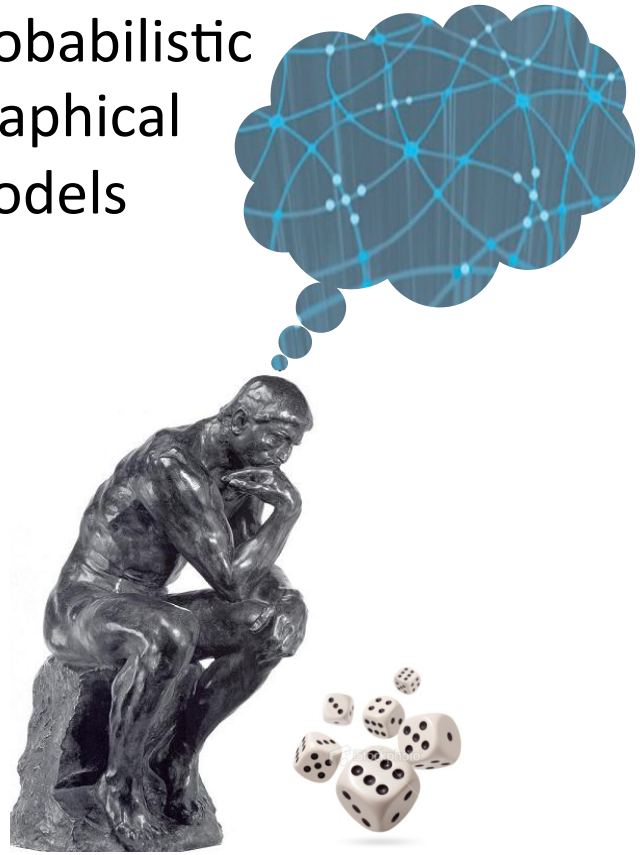


Daphne Koller

Summary

- Graph of clusters connected by sepsets
- Adjacent clusters pass information to each other about variables in sepset
  - Message from i to j summarizes everything i knows, except information obtained from j
- Algorithm may not converge
- The resulting beliefs are pseudo-marginals
- Nevertheless, very useful in practice

Probabilistic Graphical Models



Inference

Message Passing

Cluster Graph Properties

Cluster Graphs

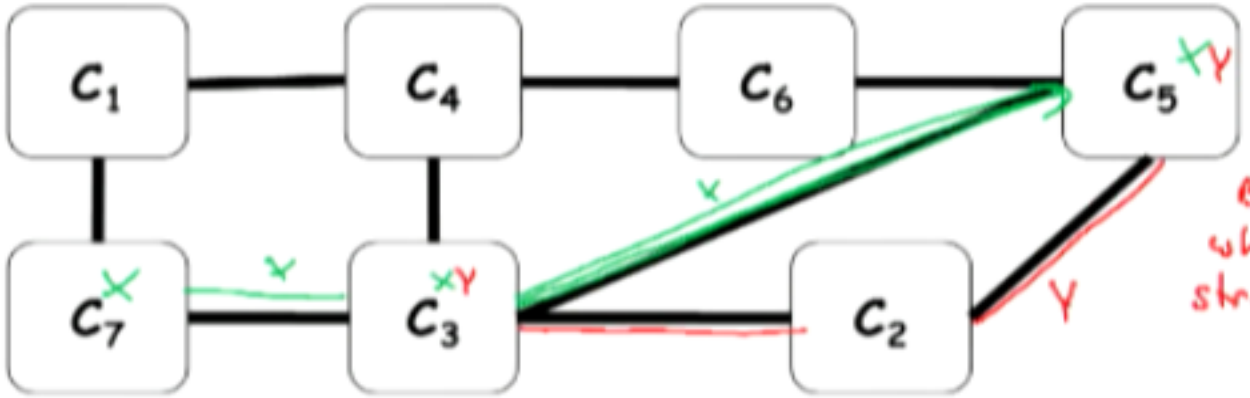
- Undirected graph such that:
  - nodes are clusters $\boldsymbol{C}_i \subseteq \{X_1,\ldots,X_n\}$
  - edge between $\boldsymbol{C}_i$ and $\boldsymbol{C}_j$ associated with sepset $\boldsymbol{S}_{i,j} \subseteq \boldsymbol{C}_i \cap \boldsymbol{C}_j$

Family Preservation

- Given set of factors Φ, we assign each ϕ_k to a cluster $C_{\alpha(k)}$ s.t. $\text{Scope}[\phi_k] \subseteq C_{\alpha(k)}$
- For each factor $\phi_k \in \Phi$, there exists a cluster C_i s.t. $\text{Scope}[\phi_k] \subseteq C_i$

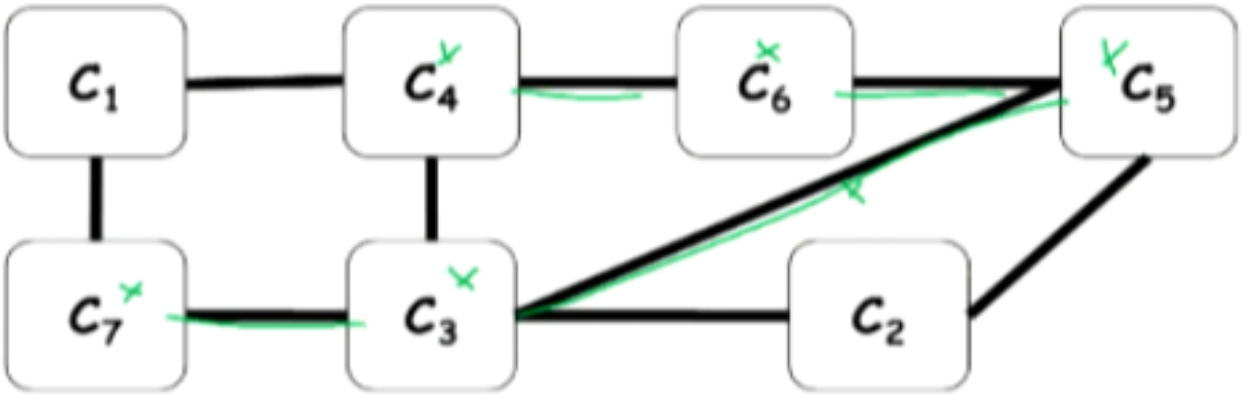
Running Intersection Property

- For each pair of clusters C_i, C_j and variable $X \in C_i \cap C_j$ there exists a unique path between C_i and C_j for which all clusters and sepsets contain X

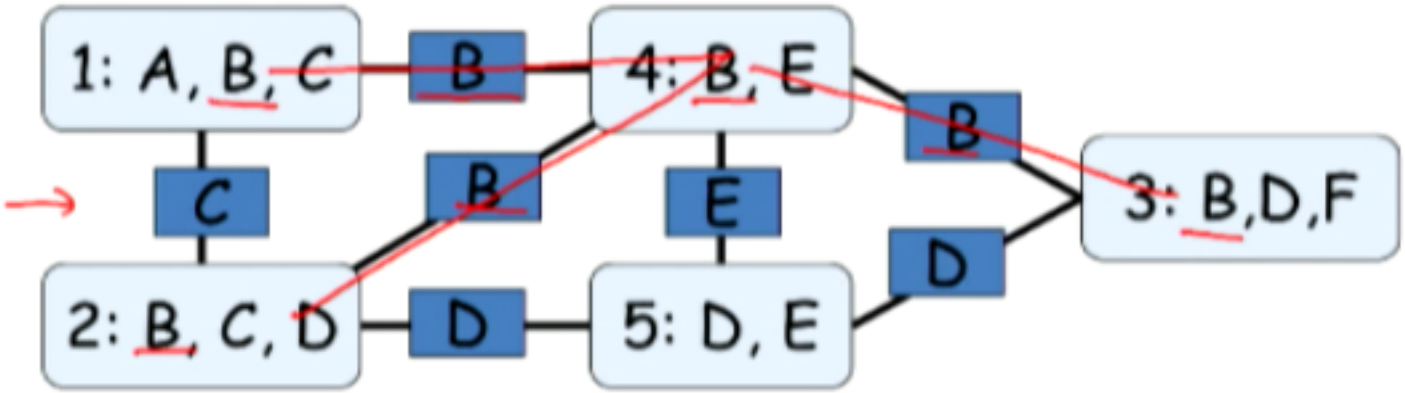


Running Intersection Property

- Equivalently: For any X, the set of clusters and sepsets containing X form a tree

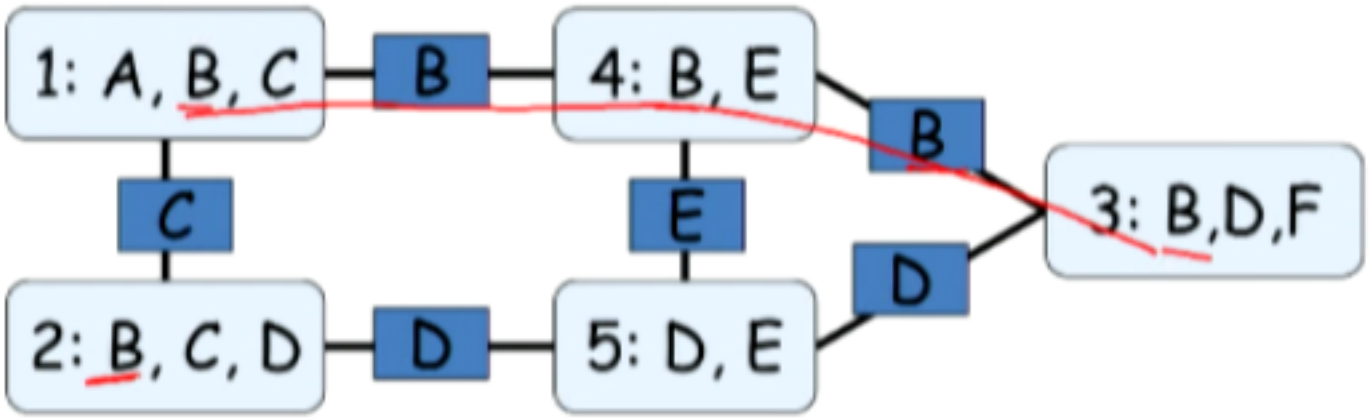


Example Cluster Graph



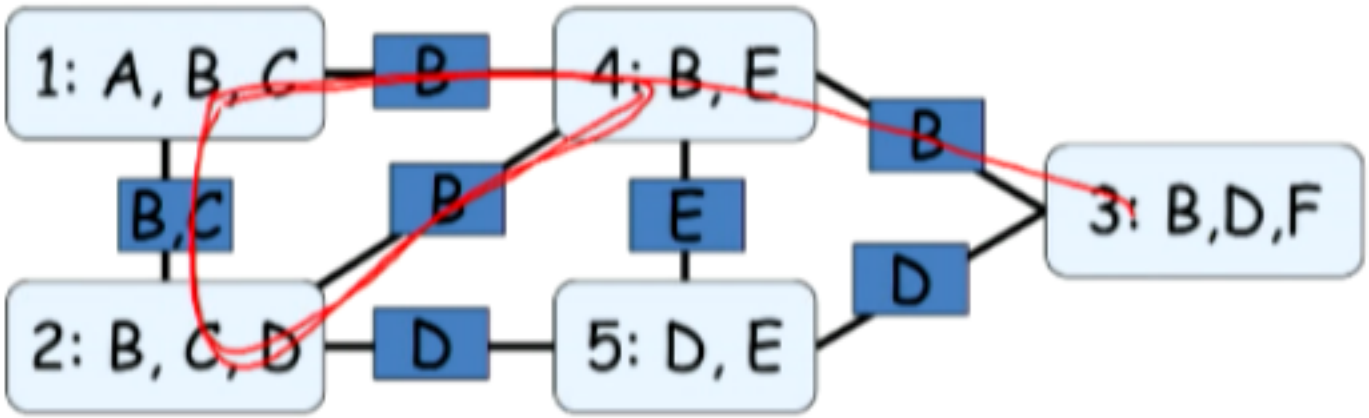
Illegal Cluster Graph I

violates e

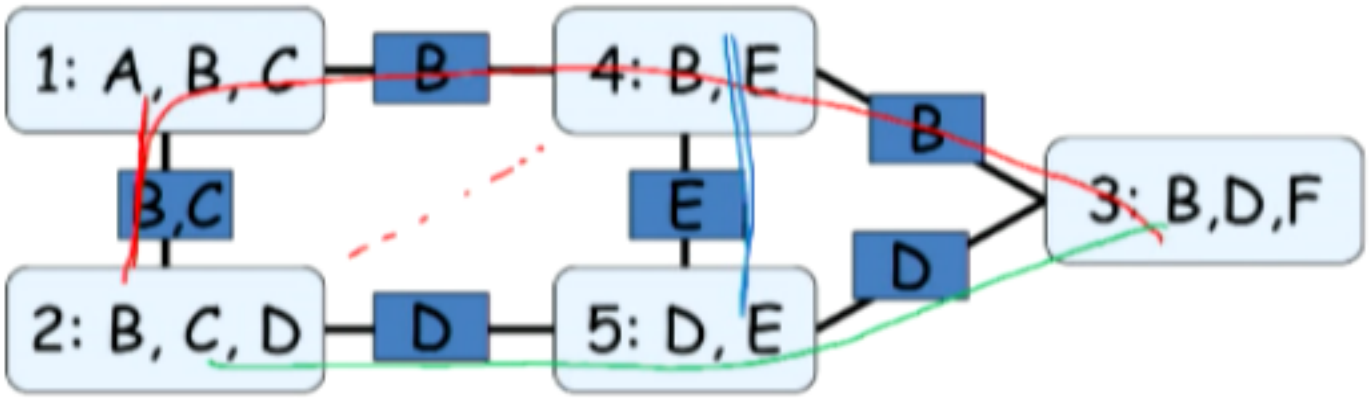


Illegal Cluster Graph II

violates uniqueness



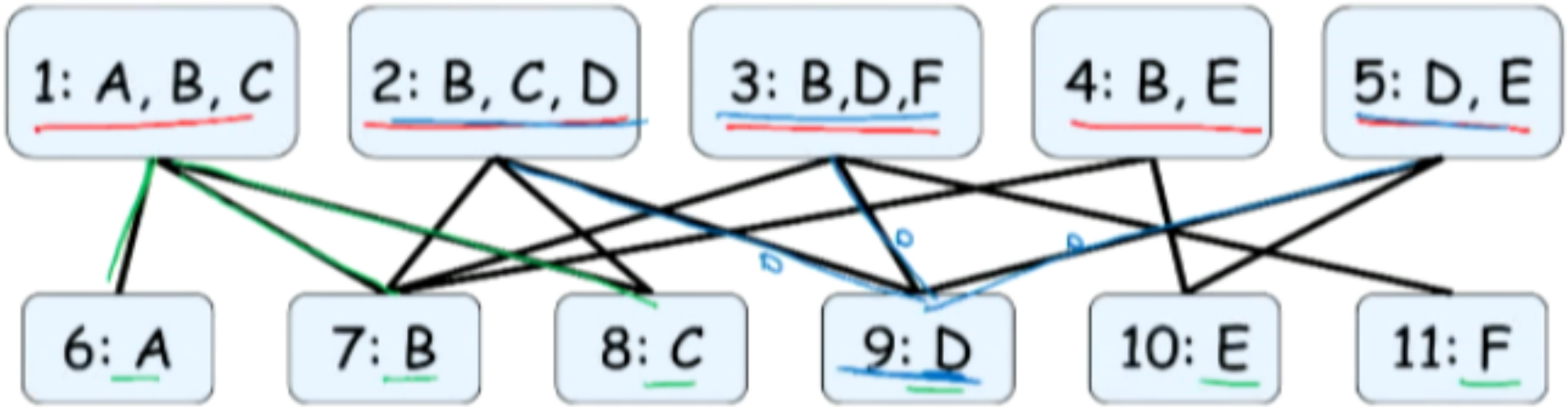
Alternative Legal Cluster Graph



Bethe Cluster Graph

big clusters = factor in Φ

- For each $\phi_k \in \Phi$, a factor cluster C_k = Scope[ϕ_k]
- For each X_i a singleton cluster $\{X_i\}$
- Edge C_k — X_i if $X_i \in C_k$



Daphne Koller

Summary

- Cluster graph must satisfy two properties
  - family preservation: allows Φ to be encoded
  - running intersection: connects all information about any variable, but without feedback loops
- Bethe cluster graph is often first default
- Richer cluster graph structures can offer different tradeoffs wrt computational cost and preservation of dependencies