Variational Approximations Tutorial

Slides by Ydo Wexler

Mean Field approximation - Reminder

- We want to approximate the **Likelihood of evidence**

  Given an evidence e such that

- The approximating distribution Q should be easy for inference – otherwise we gain nothing

.

Mean Field approximation - Reminder

$$\log P(e) = \log \sum_H P(e,h) = \log \sum_H P(e,h) \frac{Q(h)}{Q(h)}$$

$$\geq \sum_H Q(h) \log \frac{P(e,h)}{Q(h)} = -D\big(Q(h) \,\|\, P(x)\big)$$

$$D(Q \,\|\, P) = \sum_X Q(x) \log \frac{Q(x)}{P(x)}$$

.

Mean-Field Approximation

- Why are we better off computing D(Q||P) instead of computing the log-likelihood?

- Because we will choose a distribution Q the inference on which is not hard

- Mean-Field Approximation – we choose the most simple Q by:

$$Q(h) = \prod_j q_j(x_j)$$

where X_j is a single variable (node) in the network

.

Mean-Field Approximation – Setting Q

notations:

$$f_{ij} = \begin{cases} 1 & \text{If} \quad X_j \in \{X_i, pa(X_i)\} \\ 0 & \text{Otherwise} \end{cases}$$

$$D_i = \{X_i, pa(X_i)\}$$

$$y_i = P\left(X_i \mid pa(X_i)\right)$$

.

Mean-Field Approximation – Setting Q

Mean-Field Algorithm

Input: A distribution $P = \prod y_i$ over a Bayesian network, a distribution $Q = \prod_j q_j(X_j)$ i

Output: A revised set $q_j(X_j)$ such that Q is a stationary point of D(Q||P)

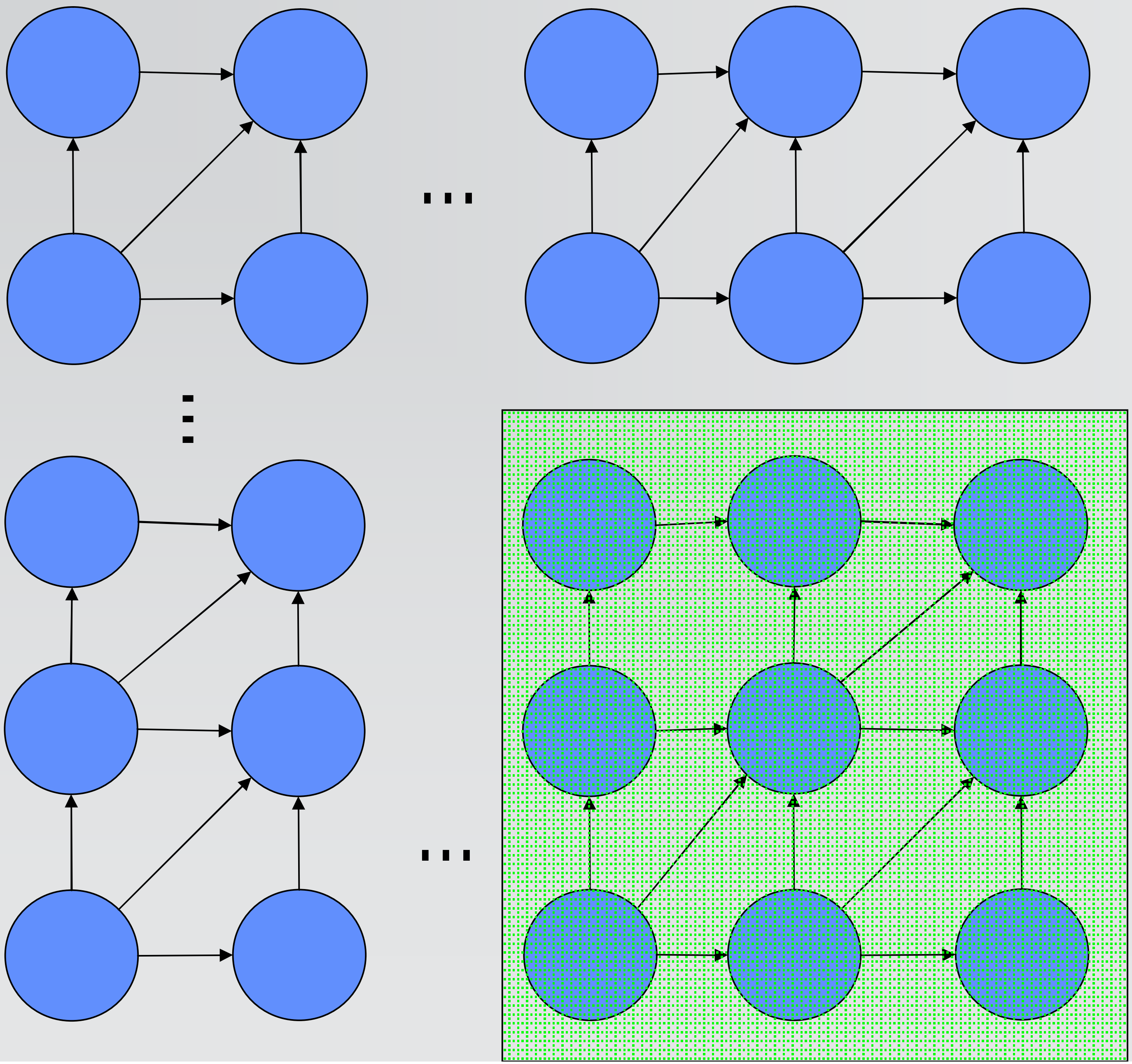
Iterate over the nodes:

$$g_j(x_j) \leftarrow \sum_{i: f_{ij=1}} \sum_{D_i \setminus X_j} \left[\prod_{\substack{k \in \{m : X_m \in D_i\} \\ k \neq j}} q_k(x_k) \right] \log y_i$$

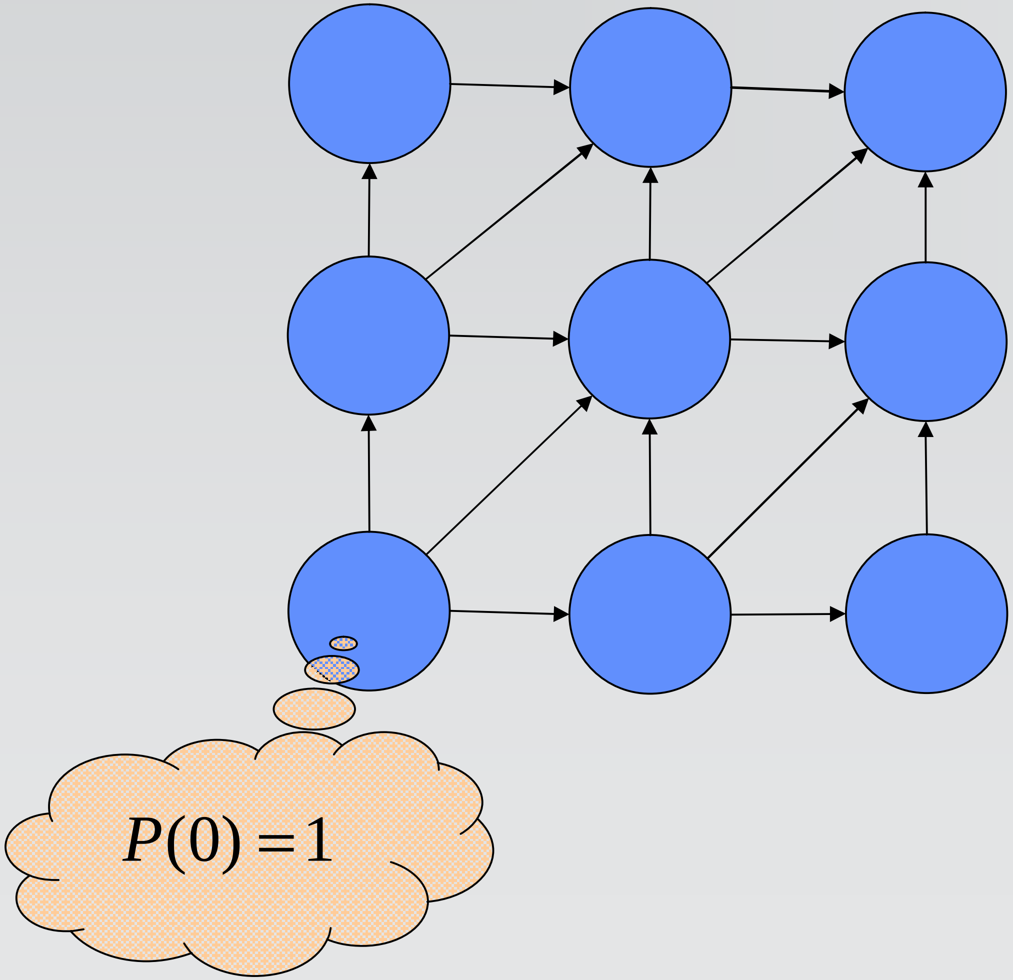
$$q_j(x_j) \leftarrow e^{g_j(x_j)}$$

Normalize q_j

Mean-Field Approximation – Example



Mean-Field Approximation – Example



$$P(a \mid a, a, a) = \frac{1}{2}$$

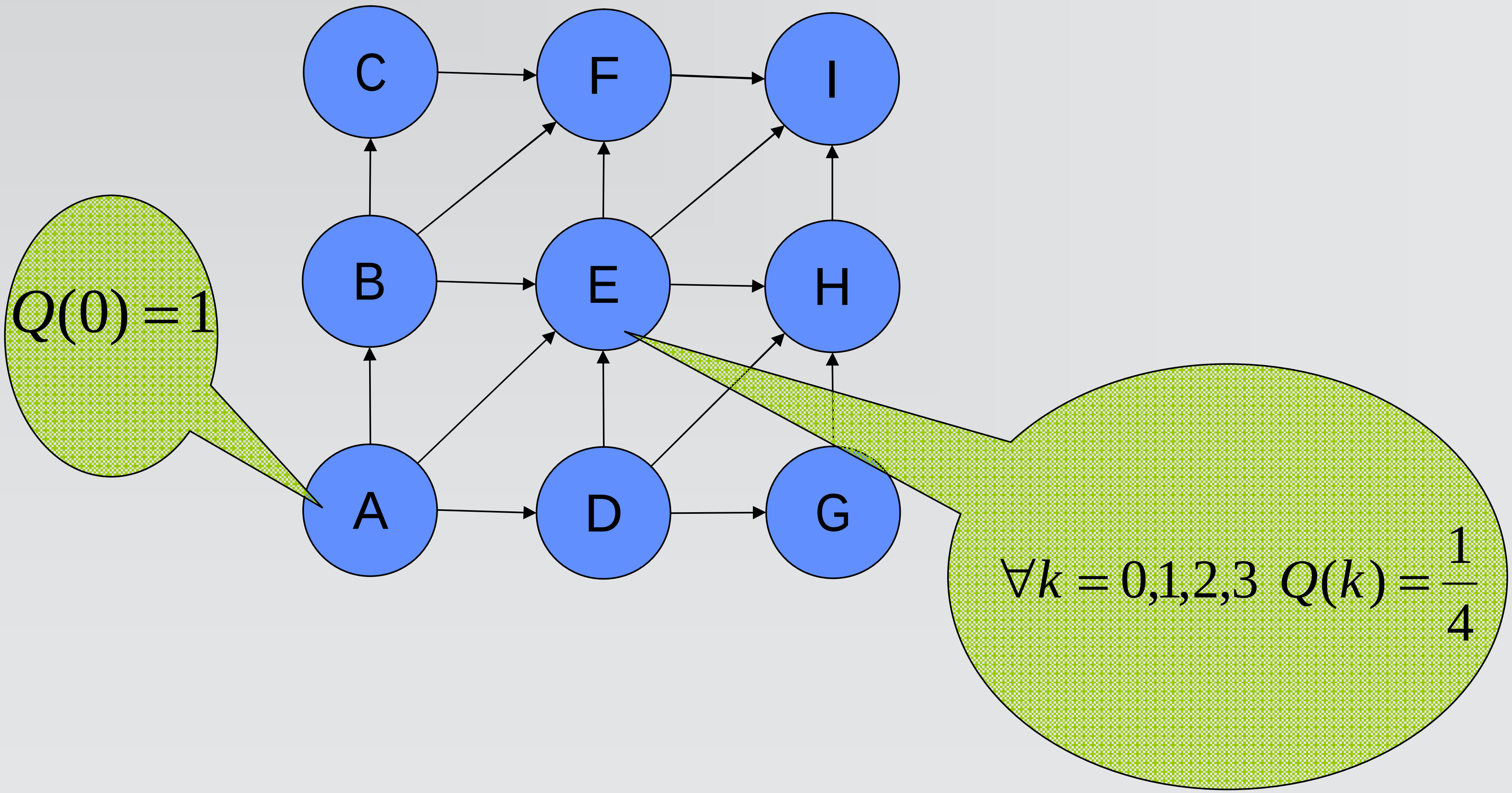
$$P(b \mid a, a, a) = \frac{1}{6}$$

$$P(\max_{a,b.c} \mid a, b, c) = \frac{1}{2}$$

$$P(d \mid a, b, c) = \frac{1}{6}$$

Mean-Field Approximation – Example

Query: likelihood I=3



Mean-Field Approximation – Example

$$D(Q \| P) = \sum_i \sum_{pa(X_i)} \sum_{X_i} \left[\prod_{j \in \{i, pa(X_i)\}} q_j(x_j) \right] \log P(x_i \mid pa(x_i))$$
$$- \sum_k \sum_{X_k} q_k(x_k) \log q_k(x_k)$$

On the edges Q=P and we know the value is 0, so we can actually ignore it and sum only on the rest

(theoretical justification – another time)

$$D(Q \| P) = \frac{1}{4} \log P(E=0 \mid 0,0,0) + \frac{1}{4} \log P(E=1 \mid 0,0,0) + \mathbf{L}$$
$$- \left(\frac{1}{4} \log \frac{1}{4} \right) \cdot 12$$

Mean-Field Approximation – Example

$$D(Q \| P) = \sum_i \sum_{pa(X_i)} \sum_{X_i} \left[\prod_{j \in \{i, pa(X_i)\}} q_j(x_j) \right] \log P(x_i \mid pa(x_i))$$
$$- \sum_k \sum_{X_k} q_k(x_k) \log q_k(x_k)$$

$$D(Q \| P) =$$

$$\frac{1}{4}\left(\log\frac{1}{2} + 3\log\frac{1}{6}\right) + 2 \cdot \left[\frac{1}{16} 2\left(\log\frac{1}{2} + 3\log\frac{1}{6}\right)\right] + \frac{1}{64}\left(37\log\frac{1}{2} + 27\log\frac{1}{6}\right)$$

$$-\left(\frac{1}{4}\log\frac{1}{4}\right) \cdot 12 \cong -2.3$$

Mean-Field Approximation – Setting Q

Mean-Field Algorithm

Input: A distribution $P = \prod y_i$ over a Bayesian network, a distribution $Q = \prod_j q_j(X_j)$ i

Output: A revised set $q_j(X_j)$ such that Q is a stationary point of D(Q||P)

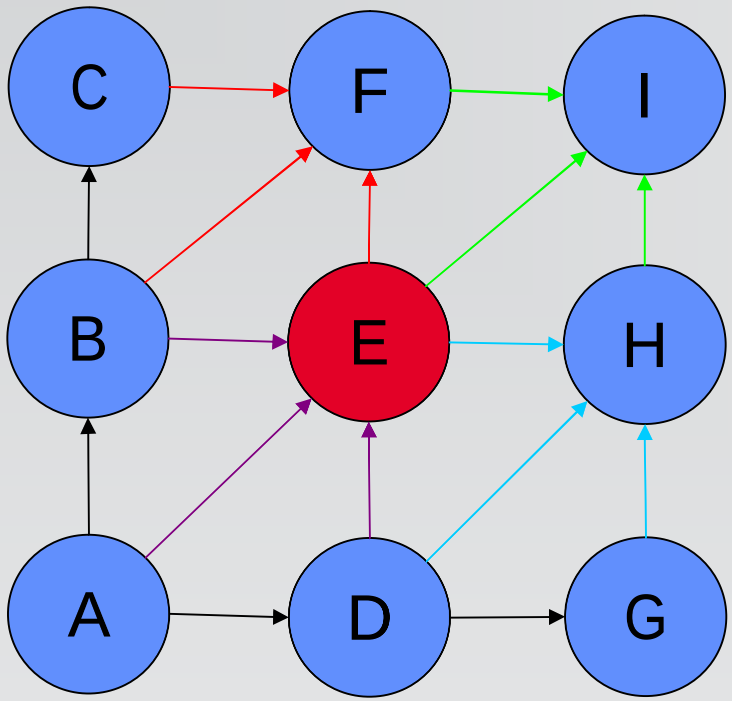
Iterate over the nodes:

$$g_j(x_j) \leftarrow \sum_{i: f_{ij=1}} \sum_{D_i \setminus X_j} \left[\prod_{\substack{k \in \{m : X_m \in D_i\} \\ k \neq j}} q_k(x_k) \right] \log y_i$$

$$q_j(x_j) \leftarrow e^{g_j(x_j)}$$

Normalize q_j

Mean-Field Approximation – Example



.

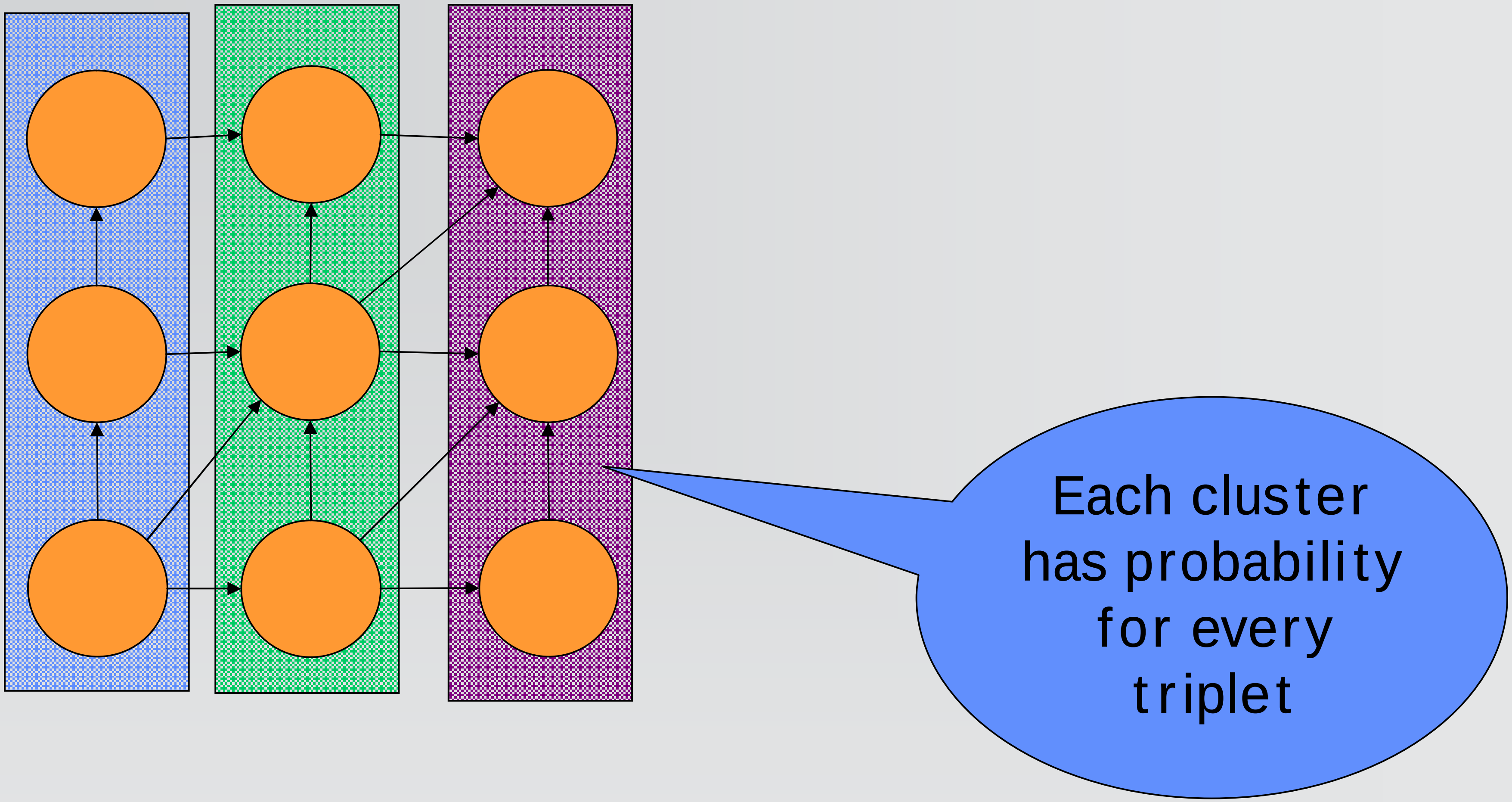
Mean-Field Approximation – Example

$$g_e(0) = \log\frac{1}{2} + \frac{1}{4}\left(\log\frac{1}{2} + 3\log\frac{1}{6}\right) + \frac{1}{4}\left(\log\frac{1}{2} + 3\log\frac{1}{6}\right) + \frac{1}{16}\left(7\log\frac{1}{2} + 9\log\frac{1}{6}\right) \cong -5.04$$

$$q_e(0) = e^{g_e(0)} \cong 0.006$$

.

Generalized Mean-Field Approximation



Home assignment – run a GMF approximation on a 3x3 example with the query P($X_{3,3}$=3)

(extra if you do 4x4 when the query is $X_{4,4}$=3)