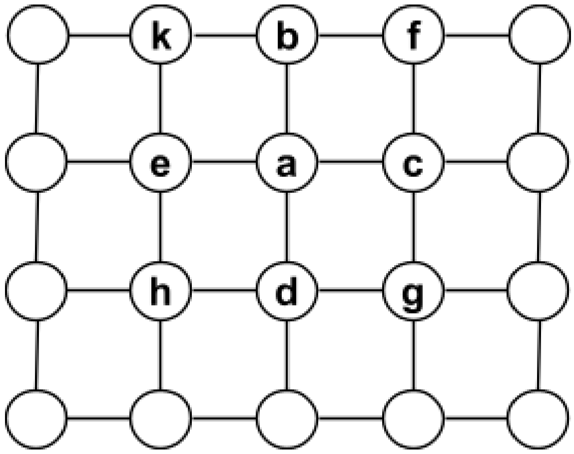
Probabilistic Graphical Models Final Exam - Spring 1400 (2021)	Instructor: B. Nasihatkon	دانشگاه صنعتی خواجه نصیرالدین طوسی K. N. TOOSI UNIVERSITY OF TECHNOLOGY

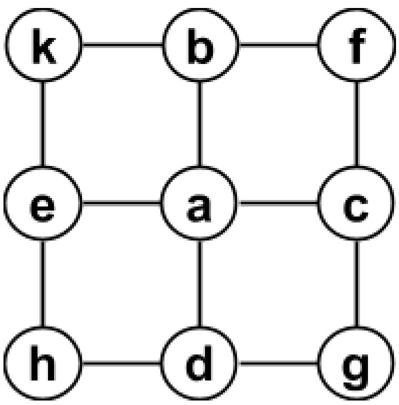
Question 1 - Message Passing on an MRF grid

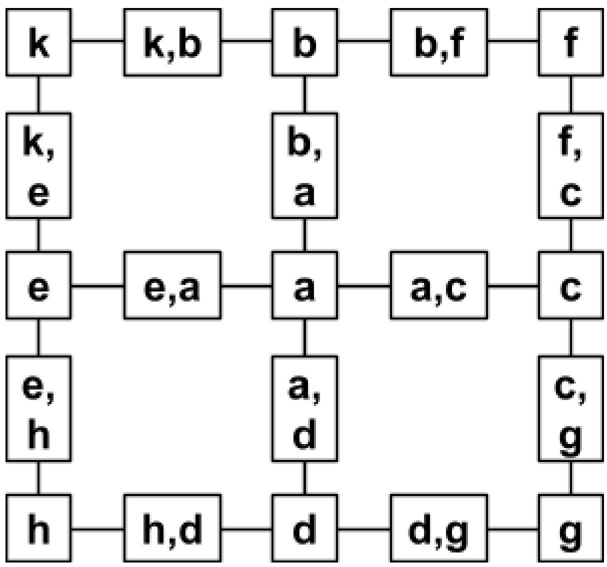
We wish to run message passing on a 4-connected MRF grid. The joint distribution is defined as



$$P(X_1, X_2, ..., X_n) = \frac{1}{Z}\exp\Big(\sum_{i \in V} f_i(X_i) + \sum_{(i,j) \in E} f_{ij}(X_i, X_j)\Big)$$

where V and E are the set of nodes and the set of edges, respectively. To perform message passing, we build a cluster graph whose clusters are the nodes and the edges of the MRF graph. The cluster graph for a 3x3 subgrid (left) is shown below on the right. **The sepsets are not shown. We assign the node potentials to the node clusters and edge potentials to edge clusters.**





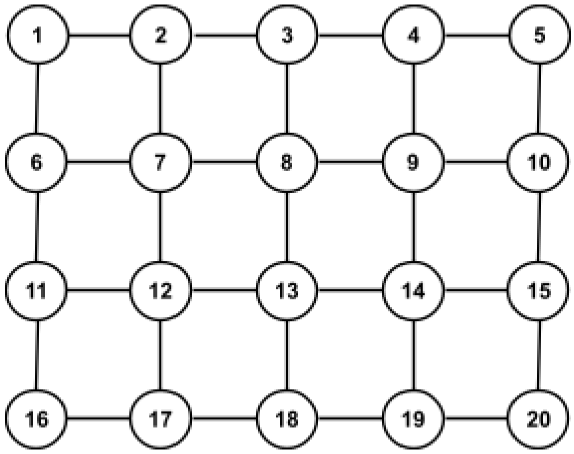
A) Show that the above cluster graph satisfies the Running Intersection Property.
B) Write down the formula for the **sum-product** messages below (in terms of $f_a(X_a)$, $f_{ac}(X_a, X_c)$, and other messages)
$\delta_{a \to (a,c)}(X_a)$
$\delta_{(a,c) \to a}(X_a)$
C) Write down the **max-sum** messages below

$\lambda_{a \to (a,c)}(X_a)$
$\lambda_{(a,c) \to a}(X_a)$

D) [Wrong] Show that if $f_{ij}(X_i, X_j) = 1(X_i = X_j)$ for all i,j, then we have
$\lambda_{(a,c) \to a}(X_a) = \max(0, \lambda_{c \to (a,c)}(X_a))$.
where $\lambda_{c \to (a,c)}(X_a)$ is the message $\lambda_{c \to (a,c)}(X_c)$evaluated at $X_c = X_a$.

Question 2 - Gibbs sampling

We intend to perform Gibbs sampling on the grid MRF defined in Question 1. Remember that the joint distribution was defined as



$$P(X_1, X_2, ..., X_n) = \frac{1}{Z}\exp\left(\sum_{i \in V} f_i(X_i) + \sum_{(i,j) \in E} f_{ij}(X_i, X_j) \right)$$

To perform a single transition

$X_1^{\,t}, X_2^{\,t}, \ ..., X_n^{\,t} \ \rightarrow \ X_1^{\,t+1}, X_2^{\,t+1}, \ ..., X_n^{\,t+1}$

of a Gibbs Markov chain each X_i^{t+1} is sampled from the distribution $R_i(X_i)$. Notice that $R_i(X_i)$ is also a function of (a subset of) the variables

$X_1^{\,t}, X_2^{\,t}, \ ..., X_n^{\,t}, \ X_1^{\,t+1}, X_2^{\,t+1}, \ ..., X_n^{\,t+1}$.

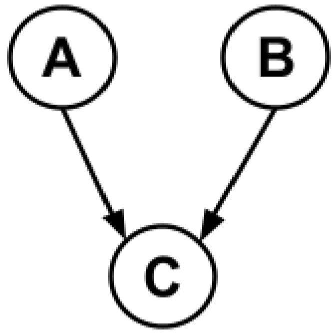
Assume that the nodes are sampled in the same order as their indices (shown in the above image), that is row-by-row, starting from top-left.

A) Derive $R_1(X_1)$.

B) Derive $R_2(X_2)$.

C) Derive $R_8(X_8)$.

Question 3 - Parameter Learning



Consider the following Bayes Net on binary variables $A, B, C \in \{0, 1\}$, with CPDs defined as:

$P(A) = \alpha A + (1 - \alpha)(1 - A),$
$P(B) = (1 + B) / 3,$
$P(C \mid A, B) = \gamma 1(C = A + B) + (1 - \gamma) 1(C \neq A + B).$

A) Write down the log-likelihood function in terms of α and γ for the following data. Simplify your answer as much as possible. **Notice that $P(C \mid A, B)$ has not been parameterized by table representation.**

a^i	b^i	c^i
0	**0**	**0**
1	**0**	**1**
1	**1**	**0**
0	**1**	**1**
0	**0**	**1**
1	**1**	**1**

B) Write down the derivatives of the log-likelihood function with respect to α and γ. Find the optimal (maximum-likelihood) values of α and γ by setting the derivatives equal to zero.