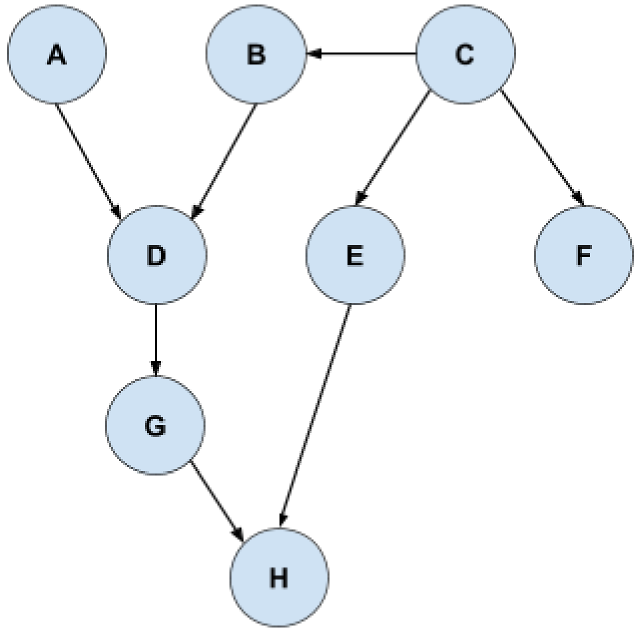
Probabilistic Graphical Models **Midterm Exam**	**Instructor:** **B. Nasihatkon**	دانشگاه صنعتی خواجه نصیرالدین طوسی K. N. TOOSI UNIVERSITY OF TECHNOLOGY
Name:	**ID:**	**Ordibehesht 1396 - May 2017**

Q1- Bayesian nets / Independence (21 points)

Consider the following Bayesian network



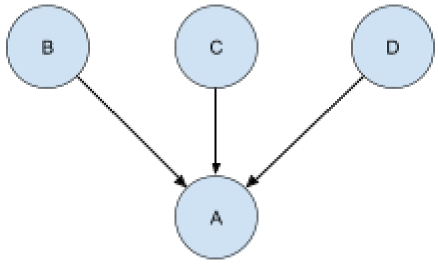
A) write the joint distribution in terms of conditional probability distributions (CPDs) **(5 points)**:

$$P(A,B,C,D,E,F,G,H) = P(A)\,P(C)\,P(B\mid C)\,P(D\mid A,B)\,P(E\mid C)\,P(F\mid C)\,P(G\mid D)\;P(H\mid G,E)$$

B) Which of the following statement are True, and which are False in general (i.e. can be false for some distribution with the above network) **(16 points)**.

	True/False			True/False
$A \perp C$	T		$A \perp B$	T
$A \perp H$	F		$H \perp C \mid E$	F
$A \perp E$	T		$A \perp B \mid D$	F
$C \perp G$	F		$H \perp F \mid C$	T
$D \perp H \mid A,\ B$	F		$C \perp A \mid H$	F
$H \perp F$	F		$F \perp A \mid H$	F
$H \perp F \mid E$	F		$F \perp A \mid B,\ H$	F
$H \perp F \mid E, G$	T		$A \perp F \mid E,\ H$	F

Q2- Bayesian nets / Context-specific independence (32 points)



Consider the conditional probability distribution $P(A \mid B, C, D)$ in the following Bayesian network, where $A, B, C, D \in \{0, 1\}$ are all binary variables. The following piece of code contains a function that receives B,C and D as arguments and returns a 2-tuple representing the probability of A=0 and A=1 given B,C,D. Therefore, the return value is equal to **(**$P(A = 0 \mid B, C, D)$, $P(A = 1 \mid B, C, D)$**)**.

```
function cpd_A_given_BCD(B,C,D)

    if B == 1
        return (.2, .8)
    else if C == 0
        return (.6, .4)
    else if D == 1
        return (.9, .1)
    else
        return (.3, .7)
```

A) Which of the following statements are True and which ones are False). Here $X \perp Y \mid Z$ means that X is independent of Y given Z **(14 Points)**.

	True/False			True/False
$A \perp B \mid C, D$	F		$A \perp D \mid B = 0, C = 0$	T
$A \perp B \mid C = 0$	F		$A \perp D \mid B = 0, C = 1$	F
$A \perp B \mid C = 1$	F		$A \perp D \mid B = 1, C = 0$	T
$A \perp C \mid B = 0$	F		$A \perp D \mid B = 1, C = 1$	T
$A \perp C \mid B = 1$	T		$B \perp C$	F
$A \perp D \mid B = 0$	F		$A \perp C \mid B = 1, D = 1$	T
$A \perp D \mid B = 1$	T		$A \perp C \mid B = 0, D = 1$	F

B) How many parameters (at least) are needed to represent a CPD $P(A \mid B, C, D)$ with the above form? Why? **(3 points)**

4 parameters. According to the above piece of code, we only need P(A=0 | B=1), P(A=0 | B=0,C=0), P(A=0 | B=0,C=1, D=1) and P(A=0 | B=0,C=1, D=0). The rest of the quantities can be obtained from these.

C) How many parameters are needed to represent a ***general*** CPD $P(A \mid B, C, D)$ for binary variables A,B,C,D (not in the above form or any other special form). Why? **(3 points)**

8 parameters are needed, namely P(A=0| B,C,D) for each combination of B,C and D. P(A=1| B,C,D) can be obtained as 1 - P(A=0| B,C,D).

D) Assume that **P(B = 0) = 0.4** and **P(C = 0) = 0.7** and **P(D = 0) = 0.2**. For the above Bayesian network compute the following probabilities. You need to write the steps towards the final answer. Do not **just** write the final solution, and do not just write the formulae **(12 points)**:

P(A = 0 | B =0, C=1, D=0) = 0.3

P(A = 0, B =0, C=1, D=0) = P(A = 0 | B =0, C=1, D=0) P(B =0, C=1, D=0)
= P(A = 0 | B =0, C=1, D=0) P(B =0) P(C=1) P(D=0)
= 0.3 * 0.4 * 0.3 * 0.2 = 0.0072

P(A = 0, B =0, C=1, D=1) = P(A = 0 | B =0, C=1, D=1) P(B =0, C=1, D=1)
= P(A = 0 | B =0, C=1, D=1) P(B =0) P(C=1) P(D=1)
= 0.9 * 0.4 * 0.3 * 0.8 = 0.0864

P(A = 0, B =0, C=1) = P(A = 0, B =0, C=1, D=0) + P(A = 0, B =0, C=1, D=1)
= 0.0072 + 0.0864 = 0.0936

P(D=0 | A = 0, B =0, C=1) = P(A = 0, B =0, C=1, D=0) / P(A = 0, B =0, C=1)
= 0.0072 / 0.0936 = 0.0769

P(A = 0 | B = 1) = 0.2 (directly from the CPD)

Extra points! (3+3 points)

P(A = 0) = P(A = 0, B =1) + P(A = 0, B = 0, C= 0) + P(A = 0, B = 0, C= 1)

P(A = 0, B =1) = P(A = 0 | B =1) P(B=1) = 0.2 * 0.6 = 0.12
P(A = 0,B = 0,C= 1) = 0.0936 (derived above)
P(A = 0,B = 0,C= 0) = P(A = 0 | B = 0,C= 0) P(B=0, C= 0) = 0.6 * P(B=0) * P(C=0) = = 0.6 * 0.4 * 0.7 = 0.168

⇒ P(A = 0) = 0.12 + 0.0936 + 0.168 = 0.3816

P(A = 0 | B = 0) = P(A = 0, B =0) / P(B=0)

P(A = 0, B =0) = P(A = 0, B =0, C=0) + P(A = 0, B =0, C=0) = 0.168 + 0.0936 (derived in previous question) = 0.2616
P(B = 0) = 0.4 (by definition)

⇒ P(A = 0 | B =0) = 0.2616 / 0.4 = 0.654

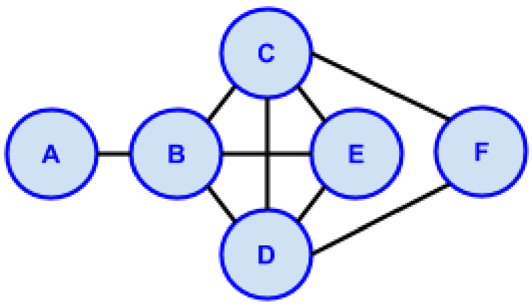
Q3- Markov Random Fields/Gibbs distribution (24 points)

Consider the following joint distribution for a Markov Random Field.
$P(A,B,C,D,E,F) = \frac{1}{Z}\phi_1(A,B)\,\phi_2(B,C,D,E)\,\phi_3(C,D,F)$

A) If $A \in \{0,1\}$, $B \in \{1,2,3\}$, $C \in \{1,2\}$, $D \in \{1,2,3,4\}$, $E \in \{0,1\}$, $F \in \{0,1\}$, how many parameters are needed in general to represent the above joint distribution? why? (assume ϕ_1, ϕ_2 and ϕ_3 do not have any specific forms) **(3 points)**
2*3 = 6 parameters for $\phi_1(A,B)$
3*2*4*2 = 48 parameters for $\phi_2(B,C,D,E)$
2*4*2 = 16 parameters for $\phi_3(C,D,F)$
A total of 6+48+16 = 70 parameters are needed

B) How many parameters were needed (at least) to parameterize an arbitrary joint distribution $P(A,B,C,D,E,F)$ (without any special structure) with the same variable domains? Why? **(2 points)**
2*3*2*4*2*2 - 1= 191 parameters, which is 3*2*4*2*2 = 192 for $P(A=a,B=b,C=c,D=d,E=e,F=f)$ for each combination of a,b,c,d,e,f, minus 1, since the sum of all probabilities add up to 1.

C) Draw the MRF graph corresponding to the above distribution. **(5 points)**



D) Which of the following statements are true and which are false (for general potential functions ϕ_1, ϕ_2 and ϕ_3)? **(14 points)**

	True/False			True/False
$A \perp B \mid C,D,E,F$	F		$F \perp E \mid C$	F
$A \perp C \mid B$	T		$A \perp D \mid B=2, C=1$	T
$A \perp B \mid C$	F		$B \perp F \mid C,\ D$	T
$A \perp F \mid C$	F		$B \perp E \mid C,D$	F
$A \perp F \mid B,C$	T		$E \perp A \mid B,C$	T
$A \perp F$	F		$E \perp F \mid B,C,D$	T
$A \perp E$	F		$F \perp A \mid B$	T

Q4- Energy Function / Graph Cuts

Consider the following energy function over the *binary* variables A, B, C, D

$E(A,B,C,D) = E_{AB}(A,B) + E_{BC}(B,C) + E_{CD}(C,D) + E_A(A) + E_B(B) + E_C(C) + E_D(D)$

where

$E_{AB}(0,0) = E_{AB}(1,1) = E_{BC}(0,0) = E_{BC}(1,1) = E_{CD}(0,0) = E_{CD}(1,1) = 0$

$E_{AB}(0,1) = E_{AB}(1,0) = 1$

$E_{BC}(0,1) = 2 \qquad E_{BC}(1,0) = 3$

$E_{CD}(0,1) = 4 \qquad E_{CD}(1,0) = 8$

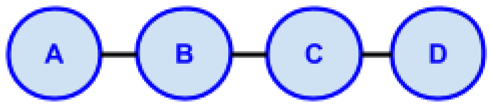
$E_A(0) = 10 \qquad E_A(1) = 12 \qquad E_B(0) = 20 \qquad E_B(1) = 22$

$E_C(0) = 40 \qquad E_C(1) = 44 \qquad E_D(0) = 80 \qquad E_D(1) = 88$

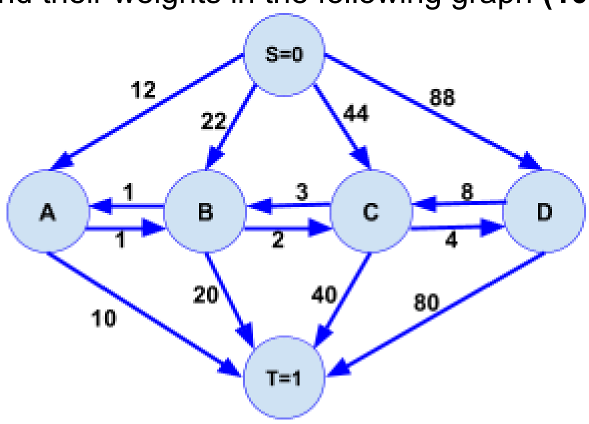
A) Write the joint distribution in terms of the (unary and binary) energy terms ($E_{AB}(A,B)$, $E_{BC}(B,C)$, $E_{CD}(C,D)$, $E_A(A)$, $E_B(B)$, $E_C(C)$ and $E_D(D)$). You do not need to compute the partition function. **(5 points)**

$$P(A,B,C,D) = \frac{1}{Z} e^{-(E_{AB}(A,B) + E_{BC}(B,C) + E_{CD}(C,D) + E_A(A) + E_B(B) + E_C(C) + E_D(D))}$$

B) Draw the corresponding MRF graph. **(3 points)**



C) Construct a min-cut/max-flow graph for the above energy function, by adding **directed** edges and their weights in the following graph **(10 points)**



D) Is $E(A,B,C,D)$ *submodular*? Why? **(5 points)**

Yes. Because for all binary energy functions E(X,Y) we have

$$E(0,1) + E(0,1) \geq E(0,0) + E(1,1)$$

$$2 = E_{AB}(0,1) + E_{AB}(1,0) \geq E_{AB}(0,0) + E_{AB}(1,1) = 0$$

$$5 = E_{BC}(0,1) + E_{BC}(1,0) \geq E_{BC}(0,0) + E_{BC}(1,1) = 0$$

$$12 = E_{CD}(0,1) + E_{CD}(1,0) \geq E_{CD}(0,0) + E_{CD}(1,1) = 0$$