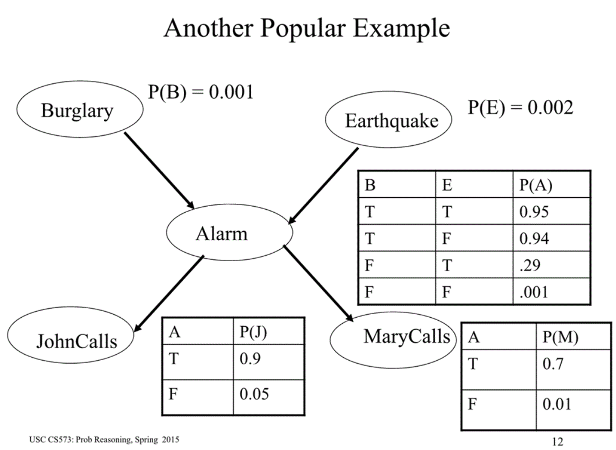
1 Question 1

Exercise 3.5 from KF book; parts b,d,f,h only
Solution:

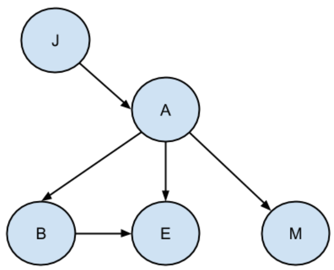
1. b) $P(d^1|t^0) > P(d^1)$. The interaction is through the active trail $D \rightarrow C \rightarrow T$. Intuitively, a good diet is more likely if we know that the person has good cholesterol levels.
2. d) $P(c^1|f^0) = P(c^1)$. No active trails between C and F given f^0. $(F \perp C)$.
3. f) $P(c^1|h^0, f^0) = P(c^1|h^0)$. No active trails. $(F \perp C|H)$.
4. h) $P(d^1|e^1, f^0, w^1) > P(d^1|e^1, f^0)$. Active trail of interaction between W and D by $D \rightarrow W$. Intuitively, a good diet is more likely given that we know the person's weight is healthy.

2 Question 2

Consider the "Alarm" network given in Lecture 3. Derive a minimal I-map for this network if we choose ordering J,A,B,M,E.



Solution:
Independencies encoded in the graph:
Notice that the edge $B \rightarrow E$ is necessary because otherwise $(B \perp E|A)$, which does not hold.



3 Question 3

We define the following properties for a set of independencies:

- **Strong Union**:

$$(\mathbf{X} \perp \mathbf{Y}|\mathbf{Z}) \implies (\mathbf{X} \perp \mathbf{Y}|\mathbf{Z}, \mathbf{W}) \tag{1}$$

- **Transitivity**:

$$\neg(\mathbf{X} \perp A|\mathbf{Z}) \wedge \neg(A \perp \mathbf{Y}|\mathbf{Z}) \implies \neg(\mathbf{X} \perp \mathbf{Y}|\mathbf{Z}) \tag{2}$$

Prove that if $\mathcal{I} = \mathcal{I}(\mathcal{H})$ for some Markov network $\mathcal{H}$, then $\mathcal{I}$ satisfies strong union and transitivity.
Solution:

Proof. **Strong Union**: Since $\mathcal{I} = \mathcal{I}(\mathcal{H})$, separation in $\mathcal{H}$ is equivalent to independence statements in distribution P. In other words, if $\mathcal{I}$ is the set of independencies that hold in distribution P, and we know $\mathcal{I}(\mathcal{H}) = \{sep_{\mathcal{H}}(\mathbf{X}; \mathbf{Y}|\mathbf{Z})\}$, we have that $\{sep_{\mathcal{H}}(\mathbf{X}; \mathbf{Y}|\mathbf{Z})\} = \mathcal{I}$. Notice that we are given $(\mathbf{X} \perp \mathbf{Y}|\mathbf{Z})$. This implies that there are no active trails between $\mathbf{X}$ and $\mathbf{Y}$ given $\mathbf{Z}$ in $\mathcal{H}$. Since the definition of separation in a Markov network is monotonic in $\mathbf{Z}$, we know that, if $sep_{\mathcal{H}}(\mathbf{X}; \mathbf{Y}|\mathbf{Z})$ holds, then $sep_{\mathcal{H}}(\mathbf{X}; \mathbf{Y}|\mathbf{Z}')$ holds for any $\mathbf{Z}' \supset \mathbf{Z}$. $\{\mathbf{Z}, \mathbf{W}\} \supset \mathbf{Z}$, and the result follows. Intuitively, observing extra evidence in Markov networks can only cause more trails to be blocked. It cannot unblock trails already blocked. Since $\mathbf{X}$ and $\mathbf{Y}$ were already separated given $\mathbf{Z}$, observing $\mathbf{W}$ as well can't un-separate them. Notice that since separation is exactly equivalent to $\mathcal{I}$, this result holds in $\mathcal{I}$. $\square$

Proof. **Transitivity**: $\neg(\mathbf{X} \perp A|\mathbf{Z})$ means that there exists an active trail between some $X \in \mathbf{X}$ and A in $\mathcal{H}$. $\neg(A \perp \mathbf{Y}|\mathbf{Z})$ means that there exists an active trail between A and some $Y \in \mathbf{Y}$ in $\mathcal{H}$. Since A is not in $\mathbf{Z}$, there must exist an active trail from some $X \in \mathbf{X}$ to some $Y \in \mathbf{Y}$ given evidence $\mathbf{Z}$ that goes through variable A, which directly implies $\neg(\mathbf{X} \perp \mathbf{Y}|\mathbf{Z})$. Since separation is again equivalent to $\mathcal{I}$, we have shown the desired transitivity result.

$\square$

4 Question 4

Show that we can represent any Gibbs distribution as a log-linear model, as defined in definition 4.15. Solution:
A Gibbs distribution can be written as:
$P(X_1, ..., X_m) = \frac{1}{Z}\Pi_{i=1}^{m}\phi_i(\mathbf{D_i})$. If we let $\epsilon_i(\mathbf{D_i}) = -ln(\phi_i(\mathbf{D_i}))$, and $\phi_i(\mathbf{D_i}) = \exp(-\epsilon_i(\mathbf{D_i}))$, we see that
$P(X_1, .., X_m) = \frac{1}{Z}\exp(-\Sigma_{i=1}^{m}\epsilon_i(\mathbf{D_i}))$. We are given that a feature is a function $f_i(\mathbf{D_i})$ is a function from $Val(\mathbf{D}) \to \mathbb{R}$. If we let $w_i = 1$ and let $f_i(\mathbf{D_i}) = \epsilon_i(\mathbf{D_i})$ for all i, we have shown that any Gibbs distribution can be converted into a log-linear model. Notice that this substitution is valid since $\epsilon_i(\mathbf{D_i})$ is also a function from $Val(\mathbf{D_i}) \to \mathbb{R}$. An alternative solution would be to create one weight/feature pair for each table entry $\mathbf{d_i}$ in each factor. The weight would be equal to $-ln(\phi_i(\mathbf{d_i})$, and the feature would be an indicator function in the form $\mathbf{1}\{\mathbf{D_i} = \mathbf{d_i}\}$. Once such indicator feature/weight pair would exist for every possible $\mathbf{d_i}$ in each factor. Thus, we could represent any Gibbs distribution as a log-linear model.

5 Question 5

This question is a simplified version of Exercise 5.13 in KF book. It starts with simplified BN20 model as in Fig. 5.1.C; for simplicity, we consider a case with only two parent and two child nodes as shown below. Now suppose that we have the observation that $F_1 = 0$, but result of F_2 are still pending. We can encode this situation in a modified BN20 network B' that has same structure as B, except F_1 is omitted from the network. Show that the posterior distribution $P(D_1, D_2|F_2, F_1 = 0)$ can be encoded using B' and specify parameters of B' in terms of parameters of B.

Hint: you can make use of proposition 5.3 without proof. For re-parameterization of B', consider changing the prior distributions of disease nodes, or the λ values on the links, or both.

Solution:

$P(D_1, D_2|F_1, F_2) = P(D_1, D_2, F_1, F_2)/P(F_1, F_2)$

$(P(F_1, F_2|D_1, D_2)P(D_1)P(D_2))/P(F_1, F_2)$

$(P(F_1 = 0|D_1, D_2)P(F2|D_1, D_2)P(D_1)P(D_2))/P(F_1, F_2)$

In B', $P(D_1, D_2|F_2') = (P(F_2'|D_1, D_2)P(D_1)P(D_2))/P(F_2')$

Let $P(F_2'|D_1, D_2) = P(F_1 = 0|D_1, D_2)P(F_2|D_1, D_2)$

$= (1 - \lambda_{1,0})(1 - \lambda_{1,1})^{d_1}(1 - \lambda_{1,2})^{d_2}P(F_2|D_1, D_2)$

$P(F_2' = 0|D_1, D_2) =$

$(1 - \lambda_{1,0})(1 - \lambda_{1,1})^{d_1}(1 - \lambda_{1,2})^{d_2}$

$(1 - \lambda_{2,0})(1 - \lambda_{2,1})^{d_1}(1 - \lambda_{2,2})^{d_2}$

$= (1 - (\lambda_{1,0} + \lambda_{2,0} - \lambda_{1,0}\lambda_{2,0}))(1 - (\lambda_{1,1} + \lambda_{2,1} - \lambda_{1,1}\lambda_{2,1}))^{d_1}$

$(1 - (\lambda_{1,2} + \lambda_{2,2} - \lambda_{1,2}\lambda_{2,2}))^{d_2}$

$\lambda_0' = \lambda_{1,0} + \lambda_{2,0} - \lambda_{1,0}\lambda_{2,0}$

$\lambda_1' = \lambda_{1,1} + \lambda_{2,1} - \lambda_{1,1}\lambda_{2,1}$

$\lambda_2' = \lambda_{2,1} + \lambda_{2,2} - \lambda_{2,1}\lambda_{2,2}$

Here we have re-parameterized the network B' in terms of the parameters of B and have shown the given posterior can be encoded in terms of B'.