① a) If W is a random variable representing the negation of X and Z is a random variable representing the negation of Y, then X is also independent of Z and Y is independent of W. So we have
$(X=0) \perp (Y=0)$, $(X=0) \perp (Y=1)$, $(X=1) \perp (Y=0)$ and if $(X=0) \perp (Y=0)$ then $(X!=0) \perp (Y!=0)$ so all 4 possibilities are independent, therefore $X \perp Y$

b) For non binary variables there are multiple possibilities in the negation, so $(X=0) \perp (Y=0)$ does not imply $(X!=0) \perp (Y!=0)$ because there is more than one possibility for $X!=0$ and $Y!=0$

② X_1 is conditionally independent of X_2 given X_3 in a distribution P if P satisfies $(X=x \perp Y=y \mid Z=z)$ for all values $x \in Val(X)$, $x_2 \in Val(X_2)$, and $x_3 \in Val(X_3)$. The independence statement is universal for all values of X_1, X_2, and X_3.
Since $(X_1 \perp X_2 \mid X_3)$ we have $P(X_1 \mid X_2, X_3) = P(X_1)$, thus
$P(X_1 \cap X_2 \mid X_3) = P(X_1 \mid X_3) P(X_2 \mid X_3)$
If we let X_3 take on all possible values we have
$P(X_1 \cap X_2 \mid X_3) = \Phi_1(X_1, X_3)\, \Phi_2(X_2, X_3)$
Letting X_1 and X_2 take all values we can write
$P(X) = \Phi_1(X_1, X_3)\, \Phi_2(X_2, X_3)$

③ Variance is the expectation of the squared difference between X and its expected value. $\Rightarrow Var[X] = E[(X - E[X])^2]$
so $Var[X] = E[X^2 - 2XE[X] + (E[X])^2]$
Now $Var[X] = E[X^2] - E[2XE[X]] + E[(E[X])^2]$
Since $E[X]$ is just a number it can be factored out so
$Var[X] = E[X^2] - 2E[X]E[X] + (E[X])^2$
$= E[X^2] - (E[X])^2$

(4) $I_P(X;Y) = E_P\left[\log \frac{P(X|Y)}{P(X)}\right]$ If X and Y are independent then $P(X|Y) = P(X)$, so $E_P[\log(1)] = 0$

If X and Y are not independent then knowing what Y is will give you more information about the value of X so $\frac{P(X|Y)}{P(X)} > 1$ and $E_P[\log(x>1)] > 0$. So $I(X;Y) \geq 0$

(5) If X and Y are independent then $P(X|Y) = P(X)$ so

$$I(X;Y) = E_P\left[\log \frac{P(X|Y)}{P(X)}\right] = E_P\left[\log \frac{P(X)}{P(X)}\right] = E_P[\log(1)] = 0$$

(6) a) B, C, D, F b) A, E, F c) Yes d) No e) No f) Yes g) Yes h) No i) No j) No, because B, C are not connected by an edge

k) Yes, because every node in $G\{D, E, F\}$ is connected by an edge and adding more nodes outside the subgraph cannot produce a clique

l) B, C, D, F, E, A