Quiz6

**1.** **(1) market basket model (1pt)**

a = {b, c, d}

b = {c, d}

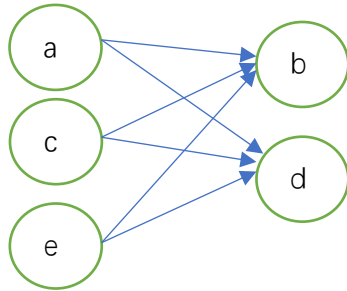c = {b, d, e, f}

d = {e, f}

e = {b, c, d}

f = {}

**(2) one example of bipartite sub-graph(1pt)**



**2.** $L = P_{wx} P_{wy} P_{xy} P_{yz} (1-P_{wz})(1-P_{xz})$

$= (P_D)^2 P_C (P_D + P_C - P_D P_C)(1-P_C)(1-\varepsilon)$ **(1pt)**

The last factor can be dropped

$P_D$ should be as large as possible

Given $P_D = 1$, the expression becomes:

$P_C (1-P_C)$ -> $P_C = 0.5$ when the expression has its maximum **(1pt)**

**3.**

Community $A$ links nodes $u, v$ independently:

$$P_A(u, v) = 1 - \exp(-F_{uA} \cdot F_{vA})$$

Then prob. at least one common $C$ links them:

$$P(u, v) = 1 - \prod_C (1 - P_C(u, v))$$
$$= 1 - \exp(-\sum_C F_{uC} \cdot F_{vC})$$
$$= 1 - \exp(-F_u \cdot F_v^T)$$ **(2pts)**

**4.**

$$x^T L x = \sum_{i,j=1}^n L_{ij} x_i x_j = \sum_{i,j=1}^n (D_{ij} - A_{ij}) x_i x_j$$

$$= \sum_i d_i x_i^2 - \sum_{(i,j) \in E} 2 x_i x_j \quad \text{for each edge } (i, j)$$

$$= \sum_{(i,j) \in E} (x_i^2 + x_j^2 - 2 x_i x_j) = \sum_{(i,j) \in E} (x_i - x_j)^2 \quad \textbf{(2pts)}$$

**5.**

**Data assumptions:**

BFR: data should be normally distributed **(0.5pts)**

CURE: data not assume any particular distribution **(0.5pts)**

**Representation of cluster:**

BFR: centroid **(0.5pts)**

CURE: a set of representatives **(0.5pts)**