Correlation Clustering Assignment

Question 1
$f(S_1) = \sum_{e \in E} w_e^-$ (all edges join different clusters so we need dissimilarity)
and
$f(S_2) = \sum_{e \in E} w_e^+$ (all edges join the same cluster so we need similarity)

Question 2
$OPT \leq \sum_{e \in E} w_e^- + \sum_{e \in E} w_e^+$ as for each edge only one of either values will be considered in the final OPT score.
$OPT \leq f(S_1) + f(S_2)$
let $f(S_{min}), f(S_{max})$ be the minimum and maximum respectively from $f(S_1), f(S_2)$.
$OPT \leq f(S_{min}) + f(S_{max})$
$OPT \leq f(S_{max}) + f(S_{max})$
$OPT \leq 2f(S_{max})$
$0.5OPT \leq f(S_{max})$
Thus algorithm is a 0.5 approximation.

Question 3
if $x_i \cdot x_j$ is 1, then they belong to the same cluster so $w_{(i,j)}^+$ should be added. We see that this holds because $w_{(i,j)}^-$ is eliminated due to multiplication by zero.
if $x_i \cdot x_j$ is 0, then they belong to different clusters so $w_{(i,j)}^-$ should be added. Again we see that this is true, because $w_{(i,j)}^+$ is eliminated for the same reason as above.

Question 4
First we notice that the two random hyperplanes are independent of each other.
The probability that a random hyperplane splits the two vectors is given as $\frac{\theta_{ij}}{\pi}$
Thus the probability they will not be splitted is $1 - \frac{\theta_{ij}}{\pi}$
Because the two hyperplanes are independent the final probability is given as the square of the above term (probability that they will not be splitted in the first phase times the probability they will not be splitted in the second phase)
Hence, $p = (1 - \frac{\theta_{ij}}{\pi})^2$

Question 5
$E[X_{ij}] = 1 \cdot p + 0 \cdot (1 - p)$
which is $E[X_{ij}] = (1 - \frac{\theta_{ij}}{\pi})^2 = g(\theta_{ij})$
$E[f(R)] = \sum_{(i,j) \in E} (w_{(i,j)}^+ \cdot E[X_{ij}]) + w_{(i,j)}^- \cdot (1 - E[X_{ij}]))$
$E[f(R)] = \sum_{(i,j) \in E} (w_{(i,j)}^+ \cdot g(\theta_{ij}) + w_{(i,j)}^- \cdot (1 - g(\theta_{ij})))$

Question 6
Using the lemma we obtain:
$E[f(R)] \geq \sum_{(i,j) \in E} (w_{(i,j)}^+ \cdot 0.75cos(\theta_{ij}) + w_{(i,j)}^- \cdot 0.75(1 - cos(\theta_{ij}))$
$E[f(R)] \geq 0.75(\sum_{(i,j) \in E} (w_{(i,j)}^+ \cdot cos(\theta_{ij}) + w_{(i,j)}^- \cdot (1 - cos(\theta_{ij})))$ **(1)**

We also know that $OPT \leq \sum_{(i,j)\in E}(w^+_{(i,j)} \cdot (x_i \cdot x_j) + w^-_{(i,j)} \cdot (1 - (x_i \cdot x_j)))$
We substitute $x_i \cdot x_j$ with its inner product which is $cos(\theta_{ij})$
$OPT \leq \sum_{(i,j)\in E}(w^+_{(i,j)} \cdot cos(\theta_{ij}) + w^-_{(i,j)} \cdot (1 - (cos(\theta_{ij})))$
Which is of course the same as below:
$\sum_{(i,j)\in E}(w^+_{(i,j)} \cdot cos(\theta_{ij}) + w^-_{(i,j)} \cdot (1 - (cos(\theta_{ij}))) \geq OPT$ **(2)**
(1)(2)
$E[f(R)] \geq 0.75OPT$
Thus the algorithm is a 0.75-Approximation