CSE547: Machine Learning for Big Data
Homework 3

Answer to Question 1(a)

proof: $w(r') = \sum_{i=1}^{n} r'_i = \sum_{i=1} \sum_{j=1} M_{ij} r_j = \sum_{j=1} \sum_{i=1} M_{ij} r_j = \sum_{j=1} r_j \sum_{i=1} M_{ij} = \sum_{j=1} r_j$
(since if there is no dead end, $\sum_j M_{ij} = 1$ for every i.)

Answer to Question 1(b)

solution: $w(r') = \sum_i r'_i = \beta \sum_i \sum_j M_{ij} r_j + (1-\beta) = \beta \sum_j r_j + (1-\beta) = \sum_j r_j = 1$ if and only if $\sum_j r_j = 1$

Answer to Question 1(c)

$r'_i = \beta(\sum_{live\, j} M_{ij} r_j + \sum_{dead\, j} r_j/n) + (1-\beta)/n.$
Thus, $w(r') = \sum_i (\beta(\sum_{live\, j} M_{ij} r_j + \sum_{dead\, j} r_j) + (1-\beta)/n)$
$= \beta(\sum_{live\, j} r_j + \sum_{dear\, j} r_j) + (1-\beta)$
$= \beta w(r) + (1-\beta) = 1$

Answer to Question 2(a)

highest:

(840, 1.0),
(155, 0.9499618624906542),
(234, 0.8986645288972261),
(389, 0.8634171101843792),
(472, 0.8632841092495216)

lowest:

(558, 0.00032860185252l5297),
(93, 0.00035135689375l6577),
(62, 0.00035314810510596274),
(424, 0.00035481538649301454),
(408, 0.00038779848719291705)

Answer to Question 2(b)

HUBBINESS:
highest:
(840, 1.0),
(155, 0.9499618624906542),
(234, 0.8986645288972261),
(389, 0.8634171101843792),
(472, 0.8632841092495216)

lowest:
(558, 0.0003286018525215297),
(93, 0.0003513568937516577),
(62, 0.00035314810510596274),
(424, 0.00035481538649301454),
(408, 0.00038779848719291705)

AUTHORITY:
highest:
(893, 1.0),
(16, 0.9635572849634397),
(799, 0.9510158161074016),
(146, 0.9246703586198444),
(473, 0.899866197360405)

lowest:
(19, 0.05608316377607618),
(135, 0.06653910487622793),
(462, 0.07544228624641901),
(24, 0.08171239406816945),
(910, 0.08571673456144878)

Answer to Question 3(a)

1. According to the definition of E[S], we have the following result:
   $2|E[S]| = 2\rho(S)|S| > \sum_v deg_S(v) = \sum_{v \in S \setminus A(S)} deg_S(v) > 2(1+\epsilon)\rho(S)|S \setminus A(S)|$,
   which indicates that $|S| > |S \setminus A(S)|(1+\epsilon) = (|S| - |A(S)|)(1+\epsilon)$
   Therefore, $A(S) \geq \frac{\epsilon}{1+\epsilon}|S|$.

2. At each iteration, we replace S by $S \setminus A(S)$ and since $|S| > |S \setminus A(S)|(1+\epsilon)$, the number of nodes becomes less than $\frac{|S|}{(1+\epsilon)}$. The number of initial number is n. Suppose after k iterations, the number of nodes is 0 and the algorithm stops, or equivalently, $\frac{|S|}{(1+\epsilon)^m} \geq 1.$, which is equivalent to $m \leq log_{1+\epsilon}n$.

Answer to Question 3(b)

1. Assume that there is one node in S*, for which $deg_{S*}(v) < \rho^*(G)$. Then we remove the node from S* and get a new set S**. $|E[S^{**}]| > |E[S^*] - \rho^*(G)$ and $|S^{**}| = |S^*| - 1$. Therefore, we have: $\rho(S^{**}) = \frac{|E[S^{**}]|}{|S^{**}|} > \frac{|E[S^*]|-\rho^*(G)}{|S^*|-1} = \frac{|E[S^*]|-\frac{|E[S^*]}{|S^*|}}{|S^*|-1} = \frac{(|S^*|-1)|E[S^*]|}{(|S^*|-1)|S^*|} = \rho(S^*) \geq \rho(S^{**})$ (since S^* has the largest density among all the subsets of G), a contradiction.

2. At first, S=V, and thus $S^* \subseteq S$. Consider the first iteration of while loop, where there exists a node $v \in S^*$ and $v \in A(S)$, before this iteration, for every node $u \in S(A)$, u is not in S^*. Thus, before this iteration, we never remove the node in S^* from S, and therefore $S^* \subseteq S$. $deg_{S^*}(v) \leq deg_S(v)$. In this iteration, since $v \in S^*$, we have: $deg_{S^*}(v) \geq \rho^*(G)$. And since $v \in A(S)$, we have: $deg_S(v) \leq 2(1+\epsilon)\rho(S)$.
   Therefore, $\rho^*(G) \leq deg_{S^*}(v) \leq deg_S(v) \leq 2(1+\epsilon)\rho(S)$

3. There must exist some iteration s.t. $v \in S^* \bigcap A(S)$. According to the above proof, we have $\rho^*(G) \leq 2(1+\epsilon)\rho(S)$, and $\rho(\tilde{S}) \geq \rho(S) \geq \frac{\rho^*(G)}{2(1+\epsilon)}$.

Answer to Question 4(a)

1. Since L = D-A, we have: $L_{ii} = d_i, L_{ij} = -A_{ij}$ for $i \neq j$.
   Let M = $\sum_{\{i,j\}\in E}(e_i - e_j)(e_i - e_j)^T$, we have $M = \sum_{\{i,j\}\in E} N_{ij}$, where $N_{ij}(i \neq j)$ is a matrix where the elements on (i,j) and (j,i) are -1, and the elements on (i,i) and (j,j) are 1. Thus, $M_{ii} = \sum_{j\ \{i,j\}\in E} 1 = d_i$ and $M_{ij}(i \neq j)$ = -1(if $\{i,j\} \in E$), 0(else) = $-A_{ij}$. Thus we have proved L = M.

2. $x^T Lx = x^T(\sum_{\{i,j\}\in E}(e_i - e_j)(e_i - e_j)^T)x = \sum_{\{i,j\}\in E}\{x^T(e_i - e_j)\}\{(e_i - e_j)^T x\} = \sum_{\{i,j\}\in E}(x_i - x_j)^2$

3. $x_S^T Lx_S = \sum_{\{i,j\}\in E}(x_S^{(i)} - x_S^{(j)})^2$.
   if $i \in S$ and $j \notin S$ or $j \in S$ and $i \notin S$, $(x_i - x_j)^2 = \frac{vol(\bar{S})}{vol(S)} + \frac{vol(S)}{vol(\bar{S})} - 2$.
   If $i \in S$ and $j \in S$ or $i \notin S$ and $j \in S$, then $(x_i - x_j)^2 = 0$.
   Thus, $x_S^T Lx_S = \sum_{\{i,j\}\in E, i\in S, j\notin S} \frac{vol(\bar{S})}{vol(S)} + \frac{vol(S)}{vol(\bar{S})} + 2 + \sum_{\{i,j\}\in E, i\notin S, j\in S} \frac{vol(\bar{S})}{vol(S)} + \frac{vol(S)}{vol(\bar{S})} + 2$
   $= cut(S)(\frac{vol(\bar{S})+vol(S)}{vol(S)} + \frac{vol(S)+vol(\bar{S})}{vol(\bar{S})}) + cut(\bar{S})(\frac{vol(\bar{S})+vol(S)}{vol(S)} + \frac{vol(\bar{S})+vol(S)}{vol(\bar{S})})$
   $= trace(D)(\frac{cut(S)}{vol(S)} + \frac{cut(\bar{S})}{vol(\bar{S})}) = trace(D)NCUT(S)$ and c = trace(D)

4. For i-th element of $x_S^T D$, if $i \in S$, then it is $\sqrt{\frac{vol(\bar{S})}{vol(S)}}d_i$; if $i \notin S$, then it is $-\sqrt{\frac{vol(S)}{vol(\bar{S})}}d_i$. We then sum all the elements of $x_S^T D$ and the sum is $\sum_{i\in S}\sqrt{\frac{vol(\bar{S})}{vol(S)}}d_i + \sum_{i\notin S} -\sqrt{\frac{vol(S)}{vol(\bar{S})}}d_i = \sqrt{vol(S)vol(\bar{S})} - \sqrt{vol(S)vol(\bar{S})} = 0$

5. $x_S^T Dx_S = \sum_{i\in S}\sqrt{\frac{vol(\bar{S})}{vol(S)}}di\sqrt{\frac{vol(\bar{S})}{vol(S)}} + \sum_{i\notin S} + \sqrt{\frac{vol(S)}{vol(\bar{S})}}d_i\sqrt{\frac{vol(S)}{vol(S)}}$
   $= \sum_{i\in S}\frac{vol(\bar{S})}{vol(S)}di + \sum_{i\notin S}\frac{vol(S)}{vol(\bar{S})}d_i$
   $= vol(S) + vol(\bar{S})$
   $= \sum d_i = 2m$

Answer to Question 4(b)

First notice that L = D-A. Le = (D-A)e = De-Ae. De is a n*1 matrix with i-th element = deg(i). Ae is alse a n*1 matrix with i-th element = $\sum_j A_{ij}$ deg(i). Thus Le = 0 = 0e, which indicates that 0 is an eigenvalue of L and e is its eigenvector. Besides, according to the second problem of part(a), we have L is a positive semidefinite matrix, and therefore all eigenvalues are nonnegative. Hence, e is the eigenvector corresponding to smallest eigenvalue of L.

Next, we let $z = D^{1/2}x$, and the optimization problem is changed to:

minimize $z^T \mathcal{L} z / z^T z$

subject to $z^T D^{1/2} e = 0, z^T z = 2m$.

Since we have proved that e is the eigenvector corresponding to the smallest eigenvalue of L, then $v_1 = D^{1/2}e$ is the eigenvector corresponding to the smallest eigenvalue of $\mathcal{L}$.

Besides, since $\mathcal{L}$ is a symmetric matrix, and suppose $v_1, v_2, ..., v_n$ is eigenvectors corresponding to eigenvalues $\mu_1 \leq \mu_2, ..., \leq \mu_n$ of $\mathcal{L}$, then we could wirte $\mathcal{L} = \sum_i \mu_i v_i v_i^T$.

At this time, the optimization problem is as follows:

minimize $z^T \mathcal{L} z / z^T z$

subject to $z^T v_1 = 0, z^T z = 2m$.

We are supposed to find z*, z*$\perp v_1$ and $\|z^*\|$ =2m, s.t $z^{*T} \mathcal{L} z^* / z^{*T} z^*$ is the smallest.

Since $v_2 \perp v_1$, and $\frac{v_2 \mathcal{L} v_2}{v_2^T v_2} = \mu_2$. Thus we have: $\mu_2 \geq \frac{z^{*T} \mathcal{L} z^*}{z^{*T} z^*}$. On the other hand, for any $z \in R^n$, $z^T z = 2m$ and $z \perp v_1$, we have $z = \sum_{i=2} a_i v_i$. Thus, $\frac{z^T \mathcal{L} z}{z^T z} = z^T(\sum_i \mu_i v_i v_i^T) z / 2m = (1/2m)(\sum_{i=2} a_i v_i^T)(\sum_k \mu_k v_k v_k^T)(\sum_{j=2} a_j v_j) = (1/2m)\sum_{i,j=2}\sum_k \mu_k a_i a_j v_i^T v_k v_k^T v_j = (1/2m)\sum_{k=2} \mu_k a_k^2$ $\geq (1/2m)\mu_2 \sum_{k=2} a_k^2 = \mu_2$(since $z^T z = 2m$ we have $\sum_i a_i^2 = 2m$).

Therefore, $\frac{z^{*T} \mathcal{L} z^*}{z^{*T} z^*} \geq \mu_2 \geq \frac{z^{*T} \mathcal{L} z^*}{z^{*T} z^*}$ which directly gives: $\mu_2 = \frac{z^{*T} \mathcal{L} z^*}{z^{*T} z^*}$.

And when z=v_2, we have:$\frac{v_2^T \mathcal{L} v_2^*}{v_2^T v_2} = \mu_2$. Thus, $z^* = v_2$ is the minimizer of the optimization problem.

Go back to the original problem, we have $x^* = D^{-1/2} v_2$ is the solution.

Answer to Question 4(c)

Q(y) $= \frac{\sum_{i,j=1}[A_{ij} - \frac{d_i d_j}{2m}]\delta(y_i, y_j)}{2m}$

$= \frac{\sum_{i,j \in S}[A_{ij} - \frac{d_i d_j}{2m}] + \sum_{i,j \notin S}[A_{ij} - \frac{d_i d_j}{2m}]}{2m}$

$= \frac{\sum_{i,j \in S} A_{ij} + \sum_{i,j \notin S} A_{ij} - \frac{1}{2m}(\sum_{i,j \in S} d_i d_j + \sum_{i,j \notin S} d_i d_j)}{2m}$

$= \frac{\sum_{i,j} A_{ij} - \sum_{i \in S, j \notin S} A_{ij} - \sum_{i \notin S, j \in S} A_{ij} - \frac{1}{2m}[(vol(S))^2 + vol(\bar{S})^2]}{2m}$

$= \frac{\sum_i d_i - cut(S) - cut(\bar{S}) - \frac{1}{2m}[(vol(S) + vol(\bar{S}))^2 - 2vol(S)vol(\bar{S})]}{2m}$

$= \frac{\sum_i d_i - cut(S) - cut(\bar{S}) - \frac{1}{2m}\sum_i [di^2 - 2vol(S)vol(\bar{S})]}{2m}$

$= \frac{2m - 2cut(S) - 2m + \frac{vol(S)vol(\bar{S})}{m}}{2m}$

$= \frac{-2cut(S) + \frac{vol(S)vol(\bar{S})}{m}}{2m}$ (Here we used the fact that $\sum_i d_i = 2m$)