# CSE547: Machine Learning for Big Data
# Homework 3

## Answer to Question 1(a)

First, note that $r'_i = M_i r$, where $M_i$ is the i-th row of $M$.

$$
\begin{aligned}
w(r) &= \sum_{i \leq n} r_i \\
w(r') &= \sum_{i \leq n} r'_i = \sum_{i \leq n} M_i r = \\
&= \sum_{i \leq n} \sum_{k \leq n} M_{ik} r_k = \sum_{k \leq n} \sum_{i \leq n} M_{ik} r_k \overset{(*)}{=} \sum_{k \leq n} r_k = w(r)
\end{aligned}
$$

At step (*), we used the fact that $\sum_{i \leq n} M_{ik} = 1$.

## Answer to Question 1(b)

$$
\begin{aligned}
w(r') &= \sum_{i\le n}(\beta\sum_{k\le n} M_{ik}r_k + \frac{(1-\beta)}{n}) = \beta\sum_{i\le n}\sum_{k\le n} M_{ik}r_k + \sum_{i\le n}\frac{(1-\beta)}{n} \\
&= \beta\sum_{k\le n}\sum_{i\le n} M_{ik}r_k + (1-\beta) = \beta\sum_{k\le n} r_k + (1-\beta) = \beta w(r) + (1-\beta)
\end{aligned}
$$

Hence with $0 < \beta < 1$,

$$
w(r) = w(r') = \beta w(r) + (1-\beta) \Leftrightarrow (1-\beta)w(r) = 1-\beta \Leftrightarrow w(r) = 1
$$

## Answer to Question 1(c)

$$r'_i = \beta\left(\sum_{j\in V\setminus D} M_{ij}r_j + \sum_{j\in D}\frac{r_j}{n}\right) + (1-\beta)\frac{1}{n}$$

Then

$$\begin{aligned}
w(r') &= \sum_{i\in V}\left(\beta\left(\sum_{j\in V\setminus D} M_{ij}r_j + \sum_{j\in D}\frac{r_j}{n}\right) + (1-\beta)\frac{1}{n}\right) \\
&= \sum_{i\in V}\beta\left(\sum_{j\in V\setminus D} M_{ij}r_j + \sum_{j\in D}\frac{r_j}{n}\right) + (1-\beta)\sum_{i\in V}\frac{1}{n} \\
&= \beta\left(\sum_{j\in V\setminus D}\sum_{i\in V} M_{ij}r_j + \sum_{j\in D}\sum_{i\in V}\frac{r_j}{n}\right) + (1-\beta) \\
&= \beta(\sum_{j\in V\setminus D} r_j + \sum_{j\in D} r_j) + (1-\beta) \\
&= \beta\,\underbrace{(\sum_{j\in V} r_j)}_{=1} + (1-\beta) = 1
\end{aligned}$$

## Answer to Question 2(a)

Five node id's with the highest Page Rank scores:

- Index = 263, Score = 0.002020291181518219
- Index = 537, Score = 0.0019433415714531497
- Index = 965, Score = 0.0019254478071662631
- Index = 243, Score = 0.001852634016241731
- Index = 285, Score = 0.0018273721700645144

Five node id's with the lowest PageRank scores:

- Index = 558, Score = 0.0003286018525215297
- Index = 93, Score = 0.0003513568937516577
- Index = 62, Score = 0.00035314810510596274
- Index = 424, Score = 0.00035481538649301454
- Index = 408, Score = 0.00038779848719291705

## Answer to Question 2(b)

Hubby
Top 5

- Index = 840, Score = 1.0
- Index = 155, Score = 0.9376453848869929
- Index = 234, Score = 0.9121023430666997
- Index = 389, Score = 0.8715685191835724
- Index = 472, Score = 0.8630602079501296

Bottom 5

- Index = 23, Score = 0.05467892100684849
- Index = 835, Score = 0.06979665267550691
- Index = 141, Score = 0.07127593852892204
- Index = 539, Score = 0.0740488002962831
- Index = 889, Score = 0.08380336242624774

Authority
Top 5

- Index = 893, Score = 1.0
- Index = 16, Score = 0.9581462737431825
- Index = 146, Score = 0.9155022452971145
- Index = 799, Score = 0.9073790082710975
- Index = 473, Score = 0.8869066892681666

Bottom 5

- Index = 135, Score = 0.08313279989016191
- Index = 19, Score = 0.09665116378042653
- Index = 408, Score = 0.10027295681923493
- Index = 910, Score = 0.1128185039343638
- Index = 93, Score = 0.11703731250174902

## Answer to Question 3(a)

**(i)** Note that due to our edges being undirected and the definition of $\deg_S(i)$, we have

$$\sum_{i \in S} \deg_S(i) = 2|E[S]| = 2|S|\frac{|E[S]|}{|S|} = 2|S|\rho(S) \tag{1}$$

At the same time, it holds

$$\begin{aligned}\sum_{i \in S} \deg_S(i) \geq \sum_{i \in S \setminus A(S)} \deg_S(i) &> \sum_{i \in S \setminus A(S)} 2(1+\varepsilon)\rho(S) \\ &= \sum_{i \in S} 2(1+\varepsilon)\rho(S) - \sum_{i \in A(S)} 2(1+\varepsilon)\rho(S) \\ &= |S|2(1+\varepsilon)\rho(S) - |A(S)|2(1+\varepsilon)\rho(S) \\ &= 2(1+\varepsilon)\rho(S)(|S| - |A(S)|)\end{aligned}$$

Using Equation (1), we get

$$\begin{aligned}& 2|S|\rho(S) > 2(1+\varepsilon)\rho(S)(|S| - |A(S)|) \\ \Leftrightarrow\ & \frac{|S|}{(1+\varepsilon)} > |S| - |A(S)| \\ \Leftrightarrow\ & \frac{-\varepsilon|S|}{(1+\varepsilon)} > -|A(S)| \\ \Leftrightarrow\ & \frac{\varepsilon|S|}{(1+\varepsilon)} < |A(S)|\end{aligned}$$

which was to show.

**(ii)** We denote $S_k$ as the set $S$ in the k-th iteration of the algorithm. Using the results from (i), we get

$$\begin{aligned}& |S_{k+1}| = |S_k| - |A(S)| \leq |S_k| - \frac{\varepsilon|S_k|}{(1+\varepsilon)} = \frac{|S_k|}{(1+\varepsilon)} \\ \Leftrightarrow\quad & \frac{|S_{k+1}|}{|S_k|} \leq \frac{1}{1+\varepsilon} \\ \Rightarrow\quad & \frac{|S_{k+1}|}{|S_0|} \leq \left(\frac{1}{1+\varepsilon}\right)^{k+1} \\ \Leftrightarrow\quad & \frac{|S_0|}{|S_{K+1}|} \leq (1+\varepsilon)^{k+1} \\ \Leftrightarrow\quad & \log_{1+\varepsilon}\left(\frac{n}{|S_{k+1}|}\right) \leq k+1\end{aligned}$$

Without loss of generality, assume that k+1 is also the number of iterations the algorithm takes. Note that for an arbitrary fixed k+1, $|S_{k+1}|$ is constant. Hence the number of steps is equal to $\log_{1+\varepsilon}(\frac{n}{|S_{k+1}|}) \in \mathcal{O}(\log_{1+\varepsilon}(n))$.

## Answer to Question 3(b)

**(i)** We give a proof by contradiction. We know by definition of $S^*$:

$$\rho^*(G) = \max_{S \subset V} \rho(S) = \rho(S^*) = \frac{|E[S^*]|}{|S^*|}$$

Further, it is easy to see that for two sets $A, B, B \subset A$, it holds

$$|A \setminus B| = |A| - |A \cap B| = |A| - |B| \tag{2}$$

Now, assume the following:

$$\exists v \in S^* : \deg_{S^*}(v) < \rho(S^*) = \frac{|E[S^*]|}{|S^*|} \tag{3}$$

Let $v' \in S^*$ be a node fullfilling Equation 3. Define $S := S^* \setminus \{v'\}$. We get

$$\begin{aligned}
\rho(S) = \frac{E[S]}{|S|} &= \frac{E[S^* \setminus \{v'\}]}{|S^*| - 1} = \frac{|\{\{u, w\} \in E | u, w \in S^*\} \setminus \{\{u, v'\} \in E | u \in S^*\}|}{|S^*| - 1} \\
&\overset{(2)}{=} \frac{|\{\{u, w\} \in E | u, w \in S^*\}| - |\{\{u, v'\} \in E | u \in S^*\}|}{|S^*| - 1} = \frac{E[S^*] - \deg_{S^*}(v')}{|S^*| - 1} \\
&\overset{(3)}{>} \frac{E[S^*] - \frac{|E[S^*]|}{|S^*|}}{|S^*| - 1} = \frac{\frac{(|S^*|-1)|E[S^*]|}{|S^*|}}{|S^*| - 1} = \frac{E[S^*]}{|S^*|} = \rho(S^*)
\end{aligned}$$

which is a contradiction to the fact that $S^*$ is the subset of $V$ with highest density. Hence (3) does not hold.

**(ii)** By definition, $v \in S^* \cap A(S)$, therefore

$$\begin{aligned}
v \in S^* &\overset{(i)}{\Rightarrow} \deg_{S^*}(v) \geq \rho^*(G) \\
v \in A(S) &\Rightarrow \deg_S(v) \leq 2(1 + \varepsilon)\rho(S)
\end{aligned}$$

Further, note that in the beginning of the algorithm $S^* \subset S = V$ and since this is the first time we have a node $v \in S^* \cap A(S)$, at this point it still holds $S^* \subset S$ (no node from $S^*$ has been removed from $S$ as of now). Therefore

$$\rho^*(G) \leq \deg_{S^*}(v) \leq \deg_S(v) \leq 2(1 + \varepsilon)\rho(S)$$

which was to show.

**(iii)** Note that we eliminate nodes from S until $S = \emptyset$, hence some iteration is guaranteed to be the first one where there exists a node $v \in S^* \cap A(S)$. We know that at that point, $\rho(S) \geq \frac{1}{2(1+\varepsilon)}\rho^*(G)$ due to **(ii)**. Now if $\rho(\tilde{S}) \geq \rho(S)$, then $\rho(\tilde{S}) \geq \frac{1}{2(1+\varepsilon)}\rho^*(G)$ is already satisfied and we are done (note that $\tilde{S}$ can never be overwritten by an $S$ with smaller density. If $\rho(\tilde{S}) < \rho(S)$, then we set $\tilde{S} = S$, hence $\rho(\tilde{S}) = \rho(S) \geq \frac{1}{2(1+\varepsilon)}\rho^*(G)$.

## Answer to Question 4(a)

1. First, note that since our graph is simple, i.e. has not self-edges, $L = D - A$ has the degree $d_i$ on the i-th diagonal and $-A_{ij}$ on the (i,j) position for $i \neq j$. For the sum over outer products, note that since our graph is simple, (a) an edge $\{i,j\}$ will only be once in $E$ and (b) there is no edge $\{i,i\}$ in E. Now let $i \in V$ be arbitrary but fixed. let $j \in V$ be arbitrary such that $\{i,j\} \in E$. Then $(e_i - e_j) = [0 \dots 0\ 1\ 0 \dots 0\ \ -1\ 0 \dots 0]$ with 1 in the i-th position and -1 in the j-th position. If we take the outer product of that vector, we will have $1 \cdot 1 = 1 = (-1) \cdot (-1)$ on the ith- and j-th diagonal of the resulting matrix. Since we sum over all edges, including the $d$ edges that originate from $i$, it holds $\left(\sum_{\{i,j\} \in E} (e_i - e_j)(e_i - e_j)^T\right)_{ii} = \deg(i) = d_{ii} = L_{ii}$. Note that at the same time, for $(e_i - e_j)(e_i - e_j)^T$, it is easy to see that the (i,j)-th and (j,i)-th position will both be $1 \cdot (-1) = -1$. Since no other edge $\{i,j\}$ exists in E, $\left(\sum_{\{i,j\} \in E} (e_i - e_j)(e_i - e_j)^T\right)_{ij} = \left((e_i - e_j)(e_i - e_j)^T\right)_{ij} = -1 = -A_{ij} = L_{ij}$.
2. Let $x \in \mathbb{R}$ be arbitrary. Then using 1., we get

$$\begin{aligned} x^T L x &= x^T (\sum_{\{i,j\} \in E} (e_i - e_j)(e_i - e_j)^T) x = \sum_{\{i,j\} \in E} x^T (e_i - e_j)(e_i - e_j)^T x \\ &= \sum_{\{i,j\} \in E} (x_i - x_j)(x_i - x_j) = \sum_{\{i,j\} \in E} (x_i - x_j)^2 \end{aligned}$$

3. Using 2., we know

$$x_S^T L x_S = \sum_{\{i,j\} \in E} (x_S^{(i)} - x_S^{(j)})^2$$

We consider different cases for where $i, j$ reside.
**Case I:** $i, j \in S$ or $i, j \notin S$.
In this case, by definition $x_S^{(i)} = x_S^{(j)}$, hence $(x_S^{(i)} - x_S^{(j)})^2 = 0$.
**Case II:** $i \in S, j \notin S$:

$$(x_S^{(i)} - x_S^{(j)})^2 = \left(\sqrt{\frac{vol(\bar{S})}{vol(S)}} - \sqrt{\frac{vol(S)}{vol(\bar{S})}}\right)^2 = \frac{vol(\bar{S})}{vol(S)} + \frac{vol(S)}{vol(\bar{S})} + 2\overbrace{\sqrt{\frac{vol(\bar{S})}{vol(S)}}\sqrt{\frac{vol(S)}{vol(\bar{S})}}}^{=1}$$

**Case III:** $i \notin S, j \in S$:

$$(x_S^{(i)} - x_S^{(j)})^2 = \left(\sqrt{\frac{vol(S)}{vol(\bar{S})}} - \sqrt{\frac{vol(\bar{S})}{vol(S)}}\right)^2 = \frac{vol(\bar{S})}{vol(S)} + \frac{vol(S)}{vol(\bar{S})} + 2\overbrace{\sqrt{\frac{vol(\bar{S})}{vol(S)}}\sqrt{\frac{vol(S)}{vol(\bar{S})}}}^{=1}$$

It follows

$$\sum_{\{i,j\}\in E}(x_S^{(i)}-x_S^{(j)})^2=\sum_{\substack{\{i,j\}\in E\\ i\oplus j\in S}}\frac{vol(\bar{S})}{vol(S)}+\frac{vol(S)}{vol(\bar{S})}+2$$

Note that the condition that either i or j (XOR) has to be in S under the sum means that we sum exactly over all elements in $cut(S)$ twice. Further, we will make use of the fact that $cut(S)=cut(\bar{S})$ and $2=\frac{vol(S)}{vol(S)}+\frac{vol(\bar{S})}{vol(\bar{S})}$

$$\begin{aligned}
&\sum_{\substack{\{i,j\}\in E\\ i\oplus j\in S}}\frac{vol(\bar{S})}{vol(S)}+\frac{vol(S)}{vol(\bar{S})}+2=2cut(S)(\frac{vol(\bar{S})}{vol(S)}+\frac{vol(S)}{vol(\bar{S})}+2)\\
=&2\left(cut(S)\frac{vol(\bar{S})}{vol(S)}+cut(S)\frac{vol(S)}{vol(\bar{S})}+cut(S)\left(\frac{vol(S)}{vol(S)}+\frac{vol(\bar{S})}{vol(\bar{S})}\right)\right)\\
=&2\left(\frac{cut(S)vol(S)+cut(S)vol(\bar{S})}{vol(S)}+\frac{cut(S)vol(S)+cut(S)vol(\bar{S})}{vol(\bar{S})}\right)\\
=&2\left((vol(S)+vol(\bar{S}))\left(\frac{cut(S)}{vol(S)}+\frac{cut(S)}{vol(\bar{S})}\right)\right)\\
&=2(vol(S)+vol(\bar{S}))\left(\frac{cut(S)}{vol(S)}+\frac{cut(\bar{S})}{vol(\bar{S})}\right)\\
&=c\cdot\text{NCUT(S)}
\end{aligned}$$

with $c=2(vol(S)+vol(\bar{S}))$
4.

$$\begin{aligned}
x_S^TDe=x_S^T[d_{11}\dots d_{nn}]^T=\sum_{i\in V}x_S^{(i)}d_{ii}&=\sum_{i\in S}x_S^{(i)}d_{ii}+\sum_{i\notin S}x_S^{(i)}d_{ii}\\
&=\sum_{i\in S}\sqrt{\frac{vol(\bar{S})}{vol(S)}}d_{ii}-\sum_{i\notin S}\sqrt{\frac{vol(S)}{vol(\bar{S})}}d_{ii}\\
&=\sqrt{\frac{vol(\bar{S})}{vol(S)}}\sum_{i\in S}d_{ii}-\sqrt{\frac{vol(S)}{vol(\bar{S})}}\sum_{i\notin S}d_{ii}\\
&=\sqrt{\frac{vol(\bar{S})}{vol(S)}}vol(S)-\sqrt{\frac{vol(S)}{vol(\bar{S})}}vol(\bar{S})\\
&=\sqrt{vol(\bar{S})vol(S)}-\sqrt{vol(S)vol(\bar{S})}=0
\end{aligned}$$

5.

$$
\begin{aligned}
x_S^T D x_S &= x_S^T [d_{11} x_S^{(1)} \dots d_{nn} x_S^{(n)}]^T = \sum_{i \in V} d_{ii} (x_S^{(i)})^2 \\
&= \sum_{i \in S} d_{ii} \frac{vol(\bar{S})}{vol(S)} + \sum_{i \notin S} d_{ii} \frac{vol(S)}{vol(\bar{S})} \\
&= vol(S) \frac{vol(\bar{S})}{vol(S)} + vol(\bar{S}) \frac{vol(S)}{vol(\bar{S})} \\
&= vol(S) + vol(\bar{S}) = \sum_{i \in V} d_{ii} = 2m
\end{aligned}
$$

## Answer to Question 4(b)

We first make a couple remarks that will be useful in the main proof. First, it is easy to see that each row $i$ of the adjacency A represents the total amount of edges pertaining to the node $i$. Hence the sum over row $i$ is equal to the degree of node $i$. Then

$$Le = (D - A)e = [d_{11}, \dots, d_{nn}]^T - [d_{11}, \dots, d_{nn}]^T = 0 \tag{4}$$

Second, note that $D - A$ is symmetric and diagonally dominant with non-negative entries (on the diagonal), making $L$ a positive semidefinite matrix.
Third, by substituting $z = D^{1/2}x$ and using the fact that $D^{1/2}$ is symmetric, the optimization problem can be rewritten as follows:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} & \frac{z^T D^{-1/2} L D^{-1/2} z}{z^T z} \\ \text{subject to} \quad & z^T D^{-1/2} De = 0 \\ & z^T D^{-1/2} D D^{-1/2} z = 2m \end{aligned}$$

Define $\mathcal{L} := D^{-1/2} L D^{-1/2}$. Fourth, note that $\mathcal{L}$ is a positive semidefinite matrix, due to $L$ being positive semidefinite:

$$v^T D^{-1/2} L D^{-1/2} v = (D^{-1/2} v)^T L (D^{-1/2} v) \geq 0$$

Hence all eigenvalues of $\mathcal{L}$ are greater equal zero. Fifth, $\mathcal{L}$ is symmetric due to $L$ being symmetric and $D^{-1/2}$ being a diagonal multiplied to $L$ from both sides.
Making use of the fact that $\mathcal{L}$ is a real symmetric matrix, we can decompose $\mathcal{L} = U \Lambda U^T$ where $\Lambda$ is a diagonal matrix containing the eigenvalues of $\mathcal{L}$ and $U$ a orthogonal matrix with the corresponding eigenvectors (Eigenvalue decomposition). Then, setting $t := U^T z$, we get

$$\frac{z^T \mathcal{L} z}{z^T z} = \frac{z^T U \Lambda U^T z}{z^T z} = \frac{t^T \Lambda t}{t^T U^T U t} = \frac{t^T \Lambda t}{t^T t} = \frac{t_1^2 \lambda_1 + \ldots + t_n^2 \lambda_n}{\sum_{i \leq n} t_i^2} \tag{5}$$

which is the weighed sum of all eigenvalues of $\mathcal{L} = D^{-1/2} L D^{-1/2}$. Naturally, the term minimizes when we put the most weight onto the smallest eigenvector that fulfills our constraints. We now show that $\lambda = 0$ is an eigenvalue of $\mathcal{L}$, with eigenvector $D^{1/2}e$, using equation (4):

$$D^{-1/2} L D^{-1/2} D^{1/2} e = D^{-1/2} L e = D^{-1/2} 0 = 0$$

Hence 0 is an Eigenvalue of $\mathcal{L}$ with a corresponding eigenvector $D^{1/2}e$. Since $\mathcal{L}$ is a PSD matrix, this is also the smallest eigenvalue of $\mathcal{L}$, hence the term that minimizes (5). However, note that the eigenvector corresponding to the smallest eigenvalue does not fulfill the first constraint on our optimization problem:

$$(D^{1/2} e)^T D^{-1/2} De = e^T D^{1/2} D^{-1/2} De = e^T De = \sum_{i \in V} d_{ii} \neq 0$$

Hence the next best candidate for minimizing the optimization problem is the eigenvector $z^*$ corresponding to the second smallest eigenvalue of $\mathcal{L}$. By definition, that vector is orthogonal to the eigenvector of the smallest eigenvalue.

$$(z^*)^T D^{1/2} e = 0 \Rightarrow (z^*)^T D^{-1/2} De = (z^*)^T D^{1/2} e = 0$$

Hence said eigenvector fulfills the first constraint of our optimization problem. Further, note that (a) $z^* \neq 0$ due to it being an eigenvector and (b) $z^*$ can be scaled to fulfill the second constraint:

$$(z^*)^T D^{-1/2} D D^{-1/2} z^* = (z^*)^T z^* = 2m$$

Hence the eigenvector $z^*$ corresponding to the second smallest eigenvalue of $\mathcal{L}$ is the minimizer of our optimization problem. In terms of the original, untransformed optimization problem, this would be $x^* := D^{-1/2} z^*$, which was to show.

## Answer to Question 4(c)

We show the equivalent equation

$$\sum_{i,j=1}^{n} \left( A_{ij} - \frac{d_i d_j}{2m} \right) \delta(y_i, y_j) = -2cut(S) + \frac{1}{m} vol(S) vol(\bar{S})$$

We get

$$\begin{aligned}
& \sum_{i,j=1}^{n} \left( A_{ij} - \frac{d_i d_j}{2m} \right) \delta(y_i, y_j) = \sum_{i,j \in S}^{n} \left( A_{ij} - \frac{d_i d_j}{2m} \right) + \sum_{i,j \in \bar{S}}^{n} \left( A_{ij} - \frac{d_i d_j}{2m} \right) \\
= \quad & \sum_{i,j \in S}^{n} A_{ij} - \sum_{i,j \in S}^{n} \frac{d_i d_j}{2m} + \sum_{i,j \in \bar{S}}^{n} A_{ij} - \sum_{i,j \in \bar{S}}^{n} \frac{d_i d_j}{2m} \\
= \quad & \sum_{i,j \in S}^{n} A_{ij} - \frac{vol(S)^2}{2m} + \sum_{i,j \in \bar{S}}^{n} A_{ij} - \frac{vol(\bar{S})^2}{2m} \\
= \quad & \sum_{i,j \in S}^{n} A_{ij} + \sum_{i,j \in \bar{S}}^{n} A_{ij} - \frac{vol(\bar{S})^2 + vol(S)^2}{2m} \\
= \quad & \sum_{i \in S}^{n} d_i - \sum_{\substack{i \in S \\ j \in \bar{S}}}^{n} A_{ij} + \sum_{i \in \bar{S}}^{n} d_i - \sum_{\substack{i \in \bar{S} \\ j \in S}}^{n} A_{ij} - \frac{vol(\bar{S})^2 + vol(S)^2}{2m} \\
= \quad & vol(S) - cut(S) + vol(\bar{S}) - cut(\bar{S}) - \frac{(vol(\bar{S}) + vol(S))^2 - 2vol(S)vol(\bar{S})}{2m} \\
= \quad & -2cut(S) + vol(S) + vol(\bar{S}) - \frac{(vol(\bar{S}) + vol(S))^2 - 2vol(S)vol(\bar{S})}{vol(S) + vol(\bar{S})} \\
= \quad & -2cut(S) - \frac{-2vol(S)vol(\bar{S})}{vol(S) + vol(\bar{S})} \\
= \quad & -2cut(S) + \frac{vol(S)vol(\bar{S})}{m}
\end{aligned}$$

which was to show.

Discussion Group:
Romain Camilleri, Andrew Wagenmaker

I acknowledge and accept the Academic Integrity clause:
Emil Azadian