# CSE 547: Machine Learning for Big Data
# Homework 3

**Academic Integrity** We take academic integrity extremely seriously. We strongly encourage students to form study groups. Students may discuss and work on homework problems in groups. However, each student must write down the solutions and the code independently. In addition, each student should write down the set of people whom they interacted with.

Discussion Group (People with whom you discussed ideas used in your answers):

*Hwai-Jin Peng*

On-line or hardcopy documents used as part of your answers:

- CSE 547 Lecture Slide (Tim Althoff, UWashington)
- Fast unfolding of communities in large networks (Vincent D. Blondel et. al.)
- A Tutorial on Spectral Clustering (Ulrike von Luxburg, CMU)
- CS 224W Lecture Note (Austin Benson, Stanford)
- CMS 139 Lecture Note (Thomas Vidick, CalTech)
- Eigenvalues of the Laplacian and their relationship to the connectedness of a graph (Anne Marsden, UChicago)
- CMSC 828L Deep Learning: Normalized Cut (David W. Jacobs, UMaryland)

I acknowledge and accept the Academic Integrity clause.

*(Signed)* ***Cheng-Yen Yang***

## Answer to Question 1(a)

If there is no dead ends in the network, we knows that $\sum_{i=1}^{n} M_{ij}$ for every node $j$, which leads us to:

$$w(\mathbf{r}') = \sum_{i=1}^{n} r'_i = \sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij} r_j = \sum_{j=1}^{n} \sum_{i=1}^{n} M_{ij} r_j = \sum_{j=1}^{n} r_j = w(\mathbf{r}). \tag{1}$$

## Answer to Question 1(b)

If there is no dead ends and a teleportation probability of $1-\beta$, we have:

$$w(\mathbf{r}') = \sum_{i=1}^{n} r'_i = \sum_{i=1}^{n} \Big(\beta\big(\sum_{j=1}^{n} M_{ij} r_j\big) + \frac{1-\beta}{n}\Big) = \beta\Big(\sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij} r_j\Big) + (1-\beta), \tag{2}$$

then by 1(a), we have $\sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij} r_j = w(\mathbf{r})$, therefore:

$$w(\mathbf{r}') = \beta w(\mathbf{r}) + (1-\beta). \tag{3}$$

Thus by given $w(\mathbf{r}) = w(\mathbf{r}') = k$, we have:

$$(1-\beta)\cdot k = (1-\beta), \tag{4}$$

so $w(\mathbf{r}) = w(\mathbf{r}')$ holds if and only if $w(\mathbf{r}) = 1$.

## Answer to Question 1(c)

We first derive $r_i'$ in terms of $\beta,\ M,\ r,\ n$ and $D$:

$$r_i' = \underbrace{\sum_{j\in D}\frac{r_j}{n}}_{dead\ nodes} + \underbrace{\sum_{j\notin D}\left(\beta M_{ij}r_j + \frac{1-\beta}{n}r_j\right)}_{live\ nodes}. \tag{5}$$

Then we have:

$$w(\mathbf{r}') = \sum_{i=1}^{n} r_i' = \sum_{i=1}^{n}\left(\sum_{j\in D}\frac{r_j}{n} + \sum_{j\notin D}\left(\beta M_{ij}r_j + \frac{1-\beta}{n}r_j\right)\right) \tag{6}$$

$$= \sum_{i=1}^{n}\sum_{j\in D}\frac{r_j}{n} + \sum_{i=1}^{n}\sum_{j\notin D}\beta M_{ij}r_j + \sum_{i=1}^{n}\sum_{j\notin D}\frac{1-\beta}{n}r_j \tag{7}$$

$$= \sum_{j\in D} r_j + \sum_{j\notin D}\beta r_j + \sum_{j\notin D}(1-\beta)r_j \tag{8}$$

$$= \sum_{j=1}^{n} r_j = w(\mathbf{r}) = 1 \tag{9}$$

## Answer to Question 2(a)

Applying the community aggregation step of the Louvain algorithm, we have:

$$Q = \sum_{C \in \mathcal{C}} \left( \frac{\sum_{j,k \in C} w_{jk}}{2m} - \left( \frac{\sum_{i \in C} d_i}{2m} \right)^2 \right). \tag{10}$$

Therefore we can derive $\Delta Q = Q_C^t - (Q_C^{t-1} + Q_i^{t-1})$ where $Q_C^t$ represents the modularity of new cluster $C$ after merging and $Q_C^{t-1}$ and $Q_i^{t-1}$ represent the modularity of the original cluster $C$ and vertex $i$:

$$Q_C^t = \frac{\sum_{in} + 2 \cdot \frac{k_{i,in}}{2}}{2m} - \left( \frac{\sum_{tot} + k_i}{2m} \right)^2 \tag{11}$$

$$Q_C^{t-1} = \frac{\sum_{in}}{2m} - \left( \frac{\sum_{tot}}{2m} \right)^2 \tag{12}$$

$$Q_i^{t-1} = 0 - \left( \frac{k_i}{2m} \right)^2 \tag{13}$$

by substituting into $\Delta Q = Q_C^t - (Q_C^{t-1} + Q_i^{t-1})$ we have:

$$\begin{aligned}
\Delta Q &= Q_C^t - (Q_C^{t-1} + Q_i^{t-1}) && (14)\\
&= \underbrace{\left[ \frac{\sum_{in} + 2 \cdot \frac{k_{i,in}}{2}}{2m} - \left( \frac{\sum_{tot} + k_i}{2m} \right)^2 \right]}_{\text{new cluster } C} - \underbrace{\left[ \frac{\sum_{in}}{2m} - \left( \frac{\sum_{tot}}{2m} \right)^2 \right]}_{\text{original cluster } C} + \underbrace{\left[ \left( \frac{k_i}{2m} \right)^2 \right]}_{\text{original node } i} && (15)\\
&= \left[ \frac{\sum_{in} + k_{i,in}}{2m} - \left( \frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[ \frac{\sum_{in}}{2m} - \left( \frac{\sum_{tot}}{2m} \right)^2 - \frac{k_i}{2m} \right)^2 \right] && (16)
\end{aligned}$$

## Answer to Question 2(b)

### (1)

By the symmetry of the ring structure, we only need to consider one of the four cliques as $\sum_{in} = \sum_{j,k \in C} w_{jk} = 12$:

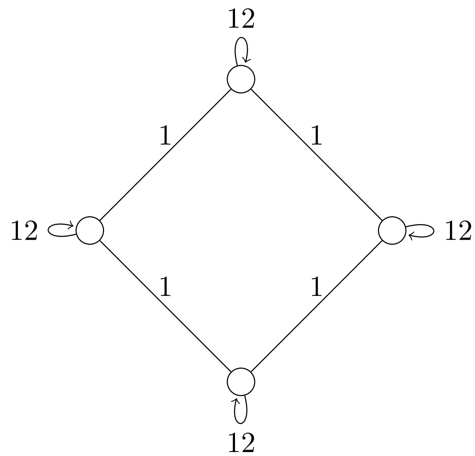


Figure 1: Graph $H$ output by the first pass of the Louvain algorithm.

### (2)

$$Q_{first} = \frac{1}{2m} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[ A_{ij} - \frac{d_i d_j}{2m} \right] \delta(c_i, c_j) = 4 \times \frac{1}{2m}(12 - \frac{12 \times 12}{2m}) = 0.6735 \tag{17}$$

### (3)

Due to the symmetry of graph $H$, two communities will be formed after the second pass of the algorithm, therefore we have the $\Delta Q$ by caculating modularity directly:

$$Q_{second} = 2 \times \frac{1}{2m}(26 - \frac{26 \times 26}{2m}) = 0.4975$$

as we can see that:

$$\Delta Q = Q_{second} - Q_{first} < 0 \tag{18}$$

therefore all node will be kept in its original communities, hence the algorithm terminates.

## Answer to Question 2(c)

As the description from the problem, density metric favors small and highly connected communities. Therefore, a better scoring function should able to rank the communities in decreasing order of the density metric and the corresponding cumulative running average of the density metric should also decreases monotonically with $k$ .

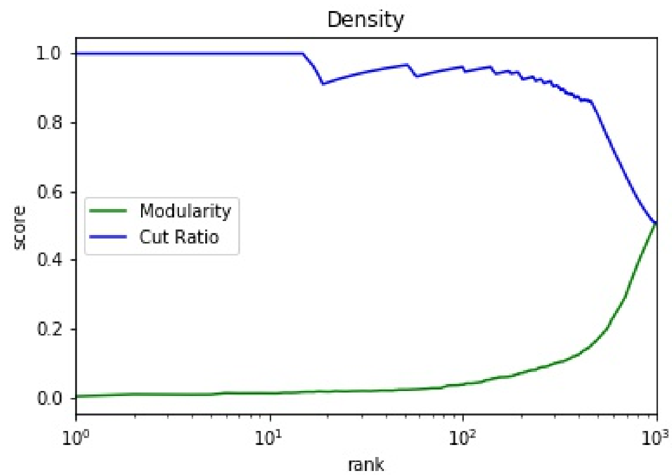


Figure 2: Density score of different scoring functions

Fig. 2 shows that the cut ratio scoring function performs better community detection in comparison to modularity scoring function in terms of the goodness metric being density function $g(S) = \frac{2m_S}{n_S(n_S-1)}$. The cumulative running average of score of cut ratio decrease monotonically as we assumed in the previous context. This also shows that modularity optimization fails to identify communities smaller than a certain scale, which is known as the resolution limit problem.

## Answer to Question 3(a)

### (1)

Let $A^{\{i,j\}}$ be the adjacency matrix of $G^{\{i,j}$ containing only one single edge $(i, j)$:

$$A_{uv}^{\{i,j\}} = \begin{cases} 1, & \text{if } \{u, v\} = \{i, j\} \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

Let $D^{\{i,j\}}$ be the diagonal matrix formed by placing the degrees$(= 1)$ of the node $i$ and $j$ along its diagonal:

$$D_{uv}^{\{i,j\}} = \begin{cases} 1, & \text{if } u = v = i \text{ or } u = v = j \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

Then by the definition of $L = D - A$, we have:

$$L = D - A = \sum_{\{i,j\} \in E} D^{\{i,j\}} - A^{\{i,j\}} = \sum_{\{i,j\} \in E} L^{\{i,j\}} \tag{21}$$

where

$$L_{uv}^{\{i,j\}} = \begin{cases} 1, & \text{if } u = v = i \text{ or } u = v = j \\ -1, & \text{if } \{u, v\} = \{i, j\} \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

which is identical to $(e_i - e_j)(e_i - e_j)^T$.
Therfore,

$$L = \sum_{\{i,j\} \in E} L^{\{i,j\}} = \sum_{\{i,j\} \in E} (e_i - e_j)(e_i - e_j)^T \tag{23}$$

# Answer to Question 3(a)

## (2)

For any arbitrary $x \in \mathcal{R}^n$:

$$
\begin{aligned}
x^T L x &= x^T (D - A) x \\
&= x^T D x - x^T A x \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j D_{ij} - \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j A_{ij} \\
&= \sum_{i=1}^{n} x_i^2 \sum_{j=1}^{n} A_{ij} - \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j A_{ij} \\
&= \sum_{\{i,j\} \in E} (x_i^2 + x_j^2) - \sum_{\{i,j\} \in E} 2 x_i x_j \\
&= \sum_{\{i,j\} \in E} (x_i - x_j)^2
\end{aligned}
$$

## Answer to Question 3(a)

### (3)

For a pre-defined vector $x_S \in \mathcal{R}^n$:

$$
\begin{aligned}
x_S^T L x_S &= \sum_{\{i,j\} \in E} \left(x_S^{(i)} - x_S^{(j)}\right)^2 && (24)\\
&= \sum_{i \in S, j \in \bar{S}} \left(x_S^{(i)} - x_S^{(j)}\right)^2 + \sum_{i \in \bar{S}, j \in S} \left(x_S^{(i)} - x_S^{(j)}\right)^2 && (25)\\
&= 2 \sum_{i \in S, j \in \bar{S}} \left(x_S^{(i)} - x_S^{(j)}\right)^2 && (26)\\
&= 2 \sum_{i \in S, j \in \bar{S}} \left(\sqrt{\frac{vol(\bar{S}}{vol(S)}} + \sqrt{\frac{vol(S)}{vol(\bar{S})}}\right)^2 && (27)\\
&= 2 \sum_{i \in S, j \in \bar{S}} \left(\frac{vol(\bar{S}}{vol(S)} + \frac{vol(S)}{vol(\bar{S})} + 2\right) && (28)\\
&= 2 \sum_{i \in S, j \in \bar{S}} \left(\frac{vol(S) + vol(\bar{S})}{vol(S)} + \frac{vol(S) + vol(\bar{S})}{vol(\bar{S})}\right) && (29)\\
&= 2\big(vol(S) + vol(\bar{S})\big) \times \sum_{i \in S, j \in \bar{S}} \left(\frac{1}{vol(S)} + \frac{1}{vol(\bar{S})}\right) && (30)\\
&= 2vol(V) \times NCUT(S) && (31)
\end{aligned}
$$

## Answer to Question 3(a)

### (4)

For a pre-defined vector $x_S \in \mathcal{R}^n$:

$$x_S^T De = \sum_{i \in V} x_S^{(i)} d_i = \sum_{i \in S} \sqrt{\frac{vol(\bar{S})}{vol(S)}} d_i - \sum_{i \in \bar{S}} \sqrt{\frac{vol(S)}{vol(\bar{S})}} d_i \tag{32}$$

by substituting $\sum_{i \in S} d_i = vol(S)$:

$$x_S^T De = \sum_{i \in S} vol(\bar{S}) vol(S) - \sum_{i \in \bar{S}} vol(S) vol(\bar{S}) = 0 \tag{33}$$

### (5)

For a pre-defined vector $x_S \in \mathcal{R}^n$:

$$x_S^T D x_S = \sum_{i \in V} \left(x_S^{(i)}\right)^2 d_i = \sum_{i \in S} \frac{vol(\bar{S})}{vol(S)} d_i - \sum_{i \in \bar{S}} \frac{vol(S)}{vol(\bar{S})} d_i \tag{34}$$

by substituting $\sum_{i \in S} d_i = vol(S)$:

$$x_S^T D = vol(\bar{S}) + vol(S) = vol(V) = 2|E| = 2m \tag{35}$$

## Answer to Question 3(b)

Using the linear transformation $z = D^{\frac{1}{2}}x$ in *Hint 1*, we can restate the problem as:

$$\underset{z \in \mathcal{R}^n}{min} \; \frac{z^T D^{-\frac{1}{2}} L D^{-\frac{1}{2}} z}{z^T z} \tag{36}$$

$$\text{subject to } z^T D^{\frac{1}{2}} e = 0 \text{ and } z^T z = 2m \tag{37}$$

Then we observe that for any $L$:

$$L \cdot e = \sum_{\{i,j\} \in E} (e_i - e_j)(e_i - e_j)^T e = 0 \cdot e = 0 \tag{38}$$

stands as $e$ is the eigenvector corresponding to the smallest eigenvalue of $L$ as *Hint 2* suggested. And for $L_{sym} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$:

$$D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \cdot D^{\frac{1}{2}} e = D^{-\frac{1}{2}} \sum_{\{i,j\} \in E} (e_i - e_j)(e_i - e_j)^T e = 0 \cdot D^{\frac{1}{2}} e = 0 \tag{39}$$

also stands as $D^{\frac{1}{2}} e$ is the eigenvector corresponding to the smallest eigenvlaue of $L_{sym}$. Both $L$ and $L_{sym}$ are positive semi-define as in part(a) we show that:

$$x^T L x = \sum_{\{i,j\} \in E} (x_i - x_j)^2 \geq 0, \, \forall x \in \mathcal{R}^n \tag{40}$$

and

$$x^T L_{sym} x = \sum_{\{i,j\} \in E} (\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}})^2 \geq 0, \, \forall x \in \mathcal{R}^n \tag{41}$$

which make them have $n$ non-negative real eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ and $\lambda_1 = 0$.
Then using the intuition of *Hint 3*, we have:

$$C = Q^T L_{sym} Q \tag{42}$$

as $Q$ being the orthonormal matrix and $C$ being the diagonal matrix containg eigenvalues of the symmetric matrix $L_{sym}$, $\lambda_1^{sym} \leq \cdots \leq \lambda_n^{sym}$ . Then by substituting $z = Qv$ into eq. (36) we have:

$$\frac{z^T L_{sym} z}{z^T z} = \frac{(Qv)^T L_{sym} (Qv)}{(Qv)^T (Qv)} = \frac{v^T C v}{v^T v} = \frac{\lambda_1^{sym} v_1^2 + \cdots + \lambda_n^{sym} v_n^2}{v_1^2 + \cdots + v_n^2} \tag{43}$$

Therefore to minimize the term, we have to pick the eigenvector of $L_{sym}$ with the smallest non-zero eigenvalue since the smallest eigenvalue 0 is correspond to the eigenvector $D^{\frac{1}{2}} e$ which will failed the constraint $z^T z = 2m$.
So the minimizer of the optimization problem is $x^* = D_2^{-\frac{1}{2}\,sym}$ where $v_2^{sym}$ is the eigenvector corresponding to the second smallest eigenvalue $\lambda_2^{sym}$.

## Answer to Question 3(c)

By definition the modularity of $y$ is:

$$Q(y) = \frac{1}{2m}\sum_{i=1}^{n}\sum_{j=1}^{n}\Big(A_{ij} - \frac{d_i d_j}{2m}\Big)\delta(y_i, y_j) \tag{44}$$

with $cut(S) = \sum_{i\in S, j\in\bar{S}} A_{ij}$ and $vol(S) = \sum_{i\in S} D_i$, we have:

$$\begin{aligned}
Q(y) &= \frac{1}{2m}\Bigg[\sum_{i,j\in S}\Big(A_{ij} - \frac{d_i d_j}{2m}\Big) + \sum_{i,j\in\bar{S}}\Big(A_{ij} - \frac{d_i d_j}{2m}\Big)\Bigg] &(45)\\
&= \frac{1}{2m}\Bigg[\Big(2m - 2cut(S)\Big) - \frac{1}{2m}\Big(vol(S)^2 + vol(\bar{S})^2\Big)\Bigg] &(46)\\
&= \frac{1}{2m}\Bigg[\Big(2m - 2cut(S)\Big) - \frac{1}{2m}\Big(\big(vol(S) + vol(\bar{S})\big)^2 - 2vol(S)vol(\bar{S})\Big)\Bigg] &(47)
\end{aligned}$$

we also know that $vol(S) + vol(\bar{S}) = 2m$, so:

$$Q(y) = \frac{1}{2m}\Big(-2cut(S) + \frac{vol(S)vol(\bar{S})}{m}\Big) \tag{48}$$