In this question we need to prove the following:

$$\Delta Q = Q_{A(fter)} - Q_{B(efore)} = \left[\frac{\Sigma_{in} + k_{i,in}}{2m} - \left(\frac{\Sigma_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\Sigma_{in}}{2m} - \left(\frac{\Sigma_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \quad (1)$$

where the modularity Q for the graph G and communities S is defined as follows:

$$Q(G,S) = \frac{1}{2m} \sum_{s \in S} \sum i, j \in s \left(A_{ij} - \frac{k_i k_j}{2m} \right) = \sum_{s \in S} \left[\frac{2|E(s)|}{2m} - \left(\frac{\sum_{v \in s} k_v}{2m} \right)^2 \right] \quad (2)$$

where, $|E(s)|$ is the weighted edges within community s. If we consider our graph with 3 communities $S = i, C, D$. i is the node we want to move to C and D is the rest of the graph (as in Fig. 2). We can calculate Q_A and Q_B using Eq. (2):

$$Q_B = \underbrace{\frac{\Sigma_{in}}{2m} - \left(\frac{k_i}{2m} \right)^2}_{C} + \underbrace{0 - \left(\frac{\Sigma_{tot}}{2m} \right)^2}_{i} \quad (3)$$

$$Q_A = \frac{\Sigma_{in} + 2 \cdot \frac{k_{i,in}}{2}}{2m} - \left(\frac{\Sigma_{tot} + k_i}{2m} \right)^2 \quad (4)$$

Note that we get Eq. (1) from Eq. (4) and (3).

Given graph G, after the first iteration of the Louvain algorithm we get graph H with the following:

- The weight of any edge between two distinct nodes in H is 1.

- The weight of any self-edge in H is 12.

- The modularity from Eq. (1) of H is $Q = 4 \cdot \left[\frac{12}{56} - \left(\frac{16}{56} \right)^2 \right] = 0.5306$.

If another iteration of this algorithm performed to produce, the graph J has the following:

- The weight of any edge between two distinct nodes in J is 2.

- The weight of any self-edge in J is 26.

- The modularity is $Q = 2 \cdot \left[\frac{26}{56} - \left(\frac{30}{56} \right)^2 \right] = 0.3546$.

Given graph G_{big}, after the first iteration of the Louvain algorithm we get graph H_{big} with the following:

- The weight of any edge between two distinct nodes in H_{big} is 1.

- The weight of any self-edge in H_{big} is 12.

- The modularity from Eq. (1) of $_{big}$ is $Q = 32 \cdot \left[\frac{12}{448} - \left(\frac{16}{448} \right)^2 \right] = 0.8163$.

If another iteration of this algorithm performed to produce, the graph J_{big} has the following:

- The weight of any edge between two distinct nodes in J_{big} is 1.

- The weight of any self-edge in J_{big} is 26.

- The modularity is $Q = 16 \cdot \left[\frac{26}{448} - \left(\frac{30}{448} \right)^2 \right] = 0.8568$.

(i) The Laplacian matrix is given by $L = D - A$ where D is the diagonal matrix of degrees and A is the adjacency matrix. Since our matrix G is undirected

$$A = \sum_{(i,j)\in E} \mathbf{e}_i\mathbf{e}_j^T + \mathbf{e}_j\mathbf{e}_i^T$$

and

$$D = \sum_{(i,j)\in E} \mathbf{e}_i\mathbf{e}_i^T + \mathbf{e}_j\mathbf{e}_j^T$$

where, $\mathbf{e}_i$ is an n-dimensional column vector with 1 at position n and 0's elsewhere. Therefore,

$$L = D - A = \sum_{(i,j)\in E} \mathbf{e}_i\mathbf{e}_i^T + \mathbf{e}_j\mathbf{e}_j^T - \mathbf{e}_i\mathbf{e}_j^T - \mathbf{e}_j\mathbf{e}_i^T = \sum_{(i,j)\in E}(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \quad (5)$$

(ii) From Eq. (5):

$$\mathbf{x}^T L\mathbf{x} = \sum_{(i,j)\in E} \mathbf{x}^T(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T\mathbf{x} = \sum_{(i,j)\in E}(x_i - x_j)(x_i - x_j) = \sum_{(i,j)\in E}(x_i - x_j)^2 \quad (6)$$

(iii) From Eq. (6):

$$\mathbf{x}^T L\mathbf{x} = \sum_{(i,j)\in E}(x_i - x_j)^2 = \sum_{\substack{(i,j)\in E \\ i\in S, j\in\bar{S}}}\left(\sqrt{\frac{\mathrm{vol}(\bar{S})}{\mathrm{vol}(S)}} + \sqrt{\frac{\mathrm{vol}(S)}{\mathrm{vol}(\bar{S})}}\right)^2 = \quad (7)$$

$$\tilde{c}\cdot\left(\frac{\mathrm{vol}(\bar{S})}{\mathrm{vol}(S)} + \frac{\mathrm{vol}(S)}{\mathrm{vol}(\bar{S})} + 2\right) = \tilde{c}\cdot\left(\frac{\mathrm{vol}(\bar{S}) + \mathrm{vol}(S)}{\mathrm{vol}(S)} + \frac{\mathrm{vol}(S) + \mathrm{vol}(\bar{S})}{\mathrm{vol}(\bar{S})}\right) =$$

$$c\cdot\left(\frac{\mathrm{cut}(S)}{\mathrm{vol}(S)} + \frac{\mathrm{cut}(\bar{S})}{\mathrm{vol}(\bar{S})}\right) = c\cdot\mathrm{NCUT}(S)$$

where $\tilde{c}$ is the number of edges in the cut and $\mathrm{cut}(S) = \mathrm{cut}(\bar{S})$.

(iv) Note that $\mathrm{vol}(S) = \sum_{i\in S} d_i$:

$$\mathbf{x}^T D\mathbf{e} = \mathbf{x}^T\mathbf{d} = \sum_i x_i d_i = \sum_{i\in S}\sqrt{\frac{\mathrm{vol}(\bar{S})}{\mathrm{vol}(S)}}d_i - \sum_{i\in\bar{S}}\sqrt{\frac{\mathrm{vol}(S)}{\mathrm{vol}(\bar{S})}}d_i = \quad (8)$$

$$\sqrt{\mathrm{vol}(\bar{S})\mathrm{vol}(S)} - \sqrt{\mathrm{vol}(S)\mathrm{vol}(\bar{S})} = 0$$

(v)

$$\mathbf{x}^T D\mathbf{x} = \sum_i x_i^2 d_i = \sum_{i\in S}\frac{\mathrm{vol}(\bar{S})}{\mathrm{vol}(S)}d_i + \sum_{i\in\bar{S}}\frac{\mathrm{vol}(S)}{\mathrm{vol}(\bar{S})}d_i = \mathrm{vol}(\bar{S}) + \mathrm{vol}(S) = \sum_i d_i = 2m \quad (9)$$

In this question we wish to solve the following problem:

$$\min_{x \in \mathbb{R}^n} \quad \frac{x^T L x}{x^T D x}$$
$$\text{s.t.} : \quad x^T D e = 0$$
$$x^T D x = 2m$$

by substitute $z = D^{1/2} x$ our new problem is:

$$\min_{x \in \mathbb{R}^n} \quad \frac{z^T \tilde{L} z}{z^T z}$$
$$\text{s.t.} : \quad z^T D^{1/2} e = 0$$
$$z^T z = 2m$$

where $\tilde{L}$ is the normalized graph Laplacian.

Using Rayleigh quotient we can minimize this problem by solving the generalized eigenvalue system,

$$\tilde{L} z = \lambda z \tag{10}$$

Note that $z_1 = D^{1/2} e$ is an eigenvector of (10) with eigenvalue 0, thus z_1 is correspondence to the smallest eigenvalue, since $\tilde{L}$ is positive semi-definite matrix since L is one (as we've seen).

Thus, the second eigenvector z_2 fulfill the first constraint since it is orthogonal to z_0 (thus, $z_2^T D^{1/2} e = z_2^T z_1 = 0$). The second constraint is fulfilled as in Eq. (9), $z_2^T z_2 = x_2^T D x_2 = 2m$.

Question 4.3, Homework 1, CS224W

Given the modularity as:

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

where y_i is 1 if $i \in S$ and -1 if $i \in \bar{S}$.

$$\sum_{1 \le i,j \le n} A_{ij} I_{y_i = y_j} = 2m - 2 \sum_{i \in S, j \in \bar{S}} A_{ij} = 2m - 2\text{cut}(S)$$

$$\sum_{1 \le i,j \le n} \frac{d_i d_j}{2m} I_{y_i = y_j} = \frac{(2m)^2}{2m} - \sum_{\substack{(i,j) \in E \\ i \in S, j \in \bar{S}}} \frac{d_i d_j}{2m} = 2m - \frac{2\text{vol}(S)\text{vol}(\bar{S})}{2m} = 2m - \frac{\text{vol}(S)\text{vol}(\bar{S})}{m}$$

thus, by applying both last equations into the first,

$$Q(y) = \frac{1}{2m} \left(-2\text{cut}(S) + \frac{\text{vol}(S)\text{vol}(\bar{S})}{m} \right)$$

Information sheet
CS224W: Machine Learning with Graphs

Assignment Submission Fill in and include this information sheet with each of your assignments. This page should be the last page of your submission. Assignments are due at 11:59pm and are always due on a Thursday. All students (SCPD and non-SCPD) must submit their homework via GradeScope (http://www.gradescope.com). Students can typeset or scan their homework. Make sure that you answer each (sub-)question on a separate page. That is, one answer per page regardless of the answer length. Students also need to upload their code on Gradescope. Put all the code for a single question into a single file and upload it.

Late Homework Policy Each student will have a total of *two* late periods. *Homework are due on Thursdays at 11:59pm PT and one late period expires on the following Monday at 11:59pm PT.* Only one late period may be used for an assignment. Any homework received after 11:59pm PT on the Monday following the homework due date will receive no credit. Once these late periods are exhausted, any assignments turned in late will receive no credit.

Honor Code We strongly encourage students to form study groups. Students may discuss and work on homework problems in groups. However, each student must write down their solutions independently, i.e., each student must understand the solution well enough in order to reconstruct it by him/herself. Students should clearly mention the names of all the other students who were part of their discussion group. Using code or solutions obtained from the web (GitHub/Google/previous year's solutions etc.) is considered an honor code violation. We check all the submissions for plagiarism. We take the honor code very seriously and expect students to do the same.

Your name: _____

Email: _____ **SUID:** _____

Discussion Group: _____

I acknowledge and accept the Honor Code.

(Signed) _____