Machine Learning for Graphs and Sequential Data Exercise Sheet 6

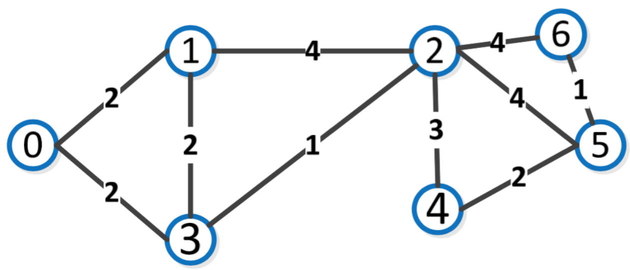
Graphs: Clustering

Problem 1: Given the graph below, find the following partitionings of the graph for $k = 2$:

a) The partitioning giving the global minimum cut

b) A partitioning approximately minimizing the ratio cut

c) A partitioning approximately minimizing the normalized cut



Problem 2: Consider minizing the ratio cut on a graph with two clusters C_1 and C_2 and N nodes in total. The indicator vector

$$f_{C_1,i} = \begin{cases} +\sqrt{\frac{|\overline{C_1}|}{|C_1|}} & \text{if } v_i \in C_1 \\ -\sqrt{\frac{|C_1|}{|\overline{C_1}|}} & \text{otherwise} \end{cases}$$

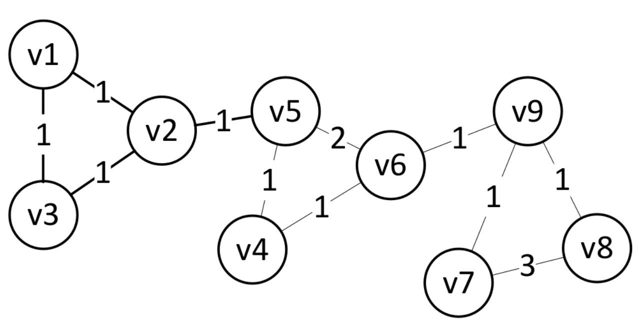
is defined as in the lecture. Prove the following three properties about $\boldsymbol{f}_{C_1}$.

a) $1^T \boldsymbol{f}_{C_1} = \sum_i f_{C_1,i} = 0$

b) $\boldsymbol{f}_{C_1}^T \boldsymbol{f}_{C_1} = \|\boldsymbol{f}_{C_1}\|_2^2 = |V|$

c) $\boldsymbol{f}_{C_1}^T L \boldsymbol{f}_{C_1} = |V| \left[\frac{\text{cut}(C_1,C_2)}{|C_1|} + \frac{\text{cut}(C_1,C_2)}{|\overline{C_1}|} \right]$

Problem 3: Answer the following questions regarding the graph below. Formulate a conjecture first and then verify it computationally in a notebook.



a) How does the first eigenvector change when increasing the weight between node $v6$ and $v9$?

b) How does the spectral embedding change?

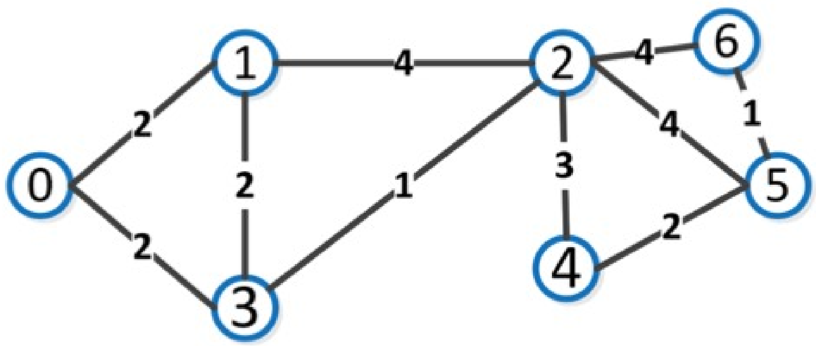
c) How does this change affect the final clustering?

Problem 1: Given the graph below, find the following partitionings of the graph for $k = 2$:

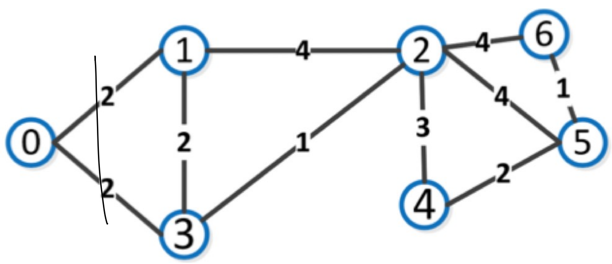
a) The partitioning giving the global minimum cut

b) A partitioning approximately minimizing the ratio cut

c) A partitioning approximately minimizing the normalized cut

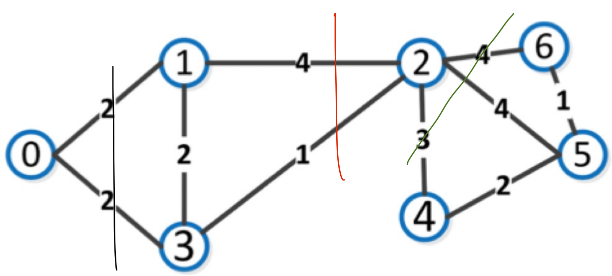


a



$$\text{cut}(c_1, c_2) = 4$$

b

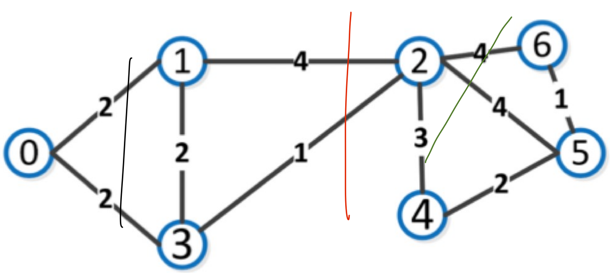


$$\frac{4}{1} + \frac{4}{6} = 4\frac{2}{3} = \frac{56}{12}$$

$$\frac{5}{3} + \frac{5}{4} = \frac{35}{12}$$

$$\frac{11}{3} + \frac{11}{4} = \frac{77}{12}$$

c



$$\frac{4}{4} + \frac{4}{42} = 1.10$$

$$\frac{5}{17} + \frac{5}{33} = 0.45.$$

$$\frac{11}{33} + \frac{11}{17} = 0.98$$

Problem 2: Consider minizing the ratio cut on a graph with two clusters C_1 and C_2 and N nodes in total. The indicator vector

$$f_{C_1,i} = \begin{cases} +\sqrt{\frac{|\overline{C_1}|}{|C_1|}} & \text{if } v_i \in C_1 \\ -\sqrt{\frac{|C_1|}{|\overline{C_1}|}} & \text{otherwise} \end{cases}$$

is defined as in the lecture. Prove the following three properties about $\boldsymbol{f}_{C_1}$.

a) $1^T \boldsymbol{f}_{C_1} = \sum_i f_{C_1,i} = 0$

b) $\boldsymbol{f}_{C_1}^T \boldsymbol{f}_{C_1} = \|\boldsymbol{f}_{C_1}\|_2^2 = |V|$

c) $\boldsymbol{f}_{C_1}^T L \boldsymbol{f}_{C_1} = |V| \left[\frac{\text{cut}(C_1, C_2)}{|C_1|} + \frac{\text{cut}(C_1, C_2)}{|\overline{C_1}|} \right]$

a) $\sum_i f_{C_1,i} = |C_1| \sqrt{\frac{|\overline{C_1}|}{|C_1|}} + |\overline{C_1}| \left(-\sqrt{\frac{|C_1|}{|\overline{C_1}|}} \right)$

$= \sqrt{|C_1||\overline{C_1}|} - \sqrt{|\overline{C_1}||C_1|}$

$= 0$

b) $\|f_{C_1}\|_2^2 = f_{C_1,1}^2 + f_{C_1,2}^2 + \dots f_{C_1,i}^2$

$= \cancel{|C_1|} \cdot \frac{|\overline{C_1}|}{\cancel{|C_1|}} + \cancel{|\overline{C_1}|} \frac{|C_1|}{\cancel{|\overline{C_1}|}}$

$= |\overline{C_1}| + |C_1|$

$= |V|$

c) $f_{C_1}^T L f_{C_1} = \frac{1}{2} \sum_{(u,v) \in E} W_{uv} (f_{C_1,u} - f_{C_1,v})^2$

$= \frac{1}{2} \Bigg[\sum_{\substack{(u,v) \in E \\ u,v \in C_1}} W_{uv} (f_{C_1,u} - f_{C_1,v})^2 + \sum_{\substack{(u,v) \in E \\ u,v \in C_2}} W_{u,v} (f_{C_1,u} - f_{C_1,v})^2$

$+ 2 \sum_{\substack{(u,v) \in E \\ u \in C_1, v \in C_2}} W_{uv} (f_{C_1,u} - f_{C_1,v})^2 \Bigg]$

if u,v in same cluster

$= \sum_{\substack{(u,v) \in E \\ u \in C_1, v \in C_2}} W_{uv} (f_{C_1,u} - f_{C_1,v})^2$

$= \underbrace{\dots W_{uv}}_{\text{cut}(C_1, C_2)} \left(\frac{|\overline{C_1}|}{|C_1|} + 2 \cdot \cancel{\sqrt{\frac{|\overline{C_1}|}{|C_1|}} \cdot \sqrt{\frac{|C_1|}{|\overline{C_1}|}}} + \frac{|C_1|}{|\overline{C_1}|} \right)$

$\frac{|\overline{C_1}|^2 + 2|C_1||\overline{C_1}| + |C_1|^2}{|C_1||\overline{C_1}|} = \frac{(|C_1| + |\overline{C_1}|)^2}{|C_1||\overline{C_1}|} = \frac{V^2}{|C_1||\overline{C_1}|}$

$= |V| \left[\frac{\text{cut}(C_1, C_2)}{|C_1|} + \frac{\text{cut}(C_1, C_2)}{|\overline{C_1}|} \right]$