Machine Learning for Graphs and Sequential Data Exercise Sheet 6

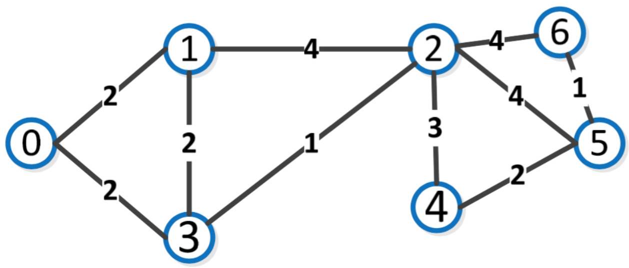
Graphs: Clustering

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

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

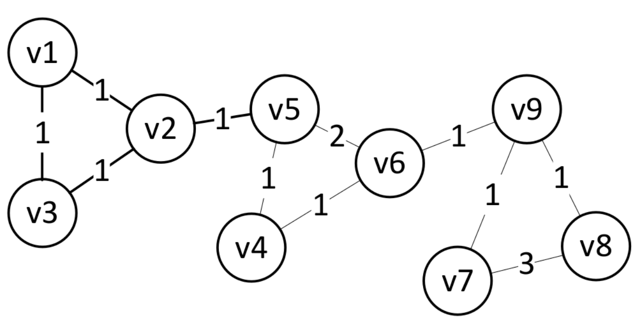
is defined as in the lecture. Prove the following three properties about $\boldsymbol{f}_{C_1}$.

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.

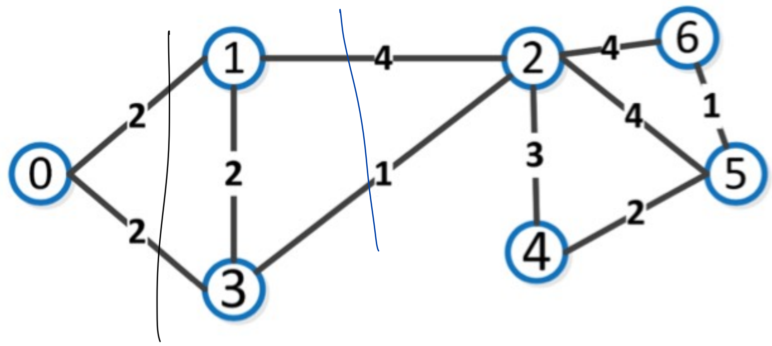


Problem 1: Given the graph below, find the following partitionings of the graph for $k = 2$:

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)

$Cut(C_1, C_2) = \sum_{V_1 \in C_1, V_2 \in C_2} W(V_1, V_2)$ is minimized

Problem : Tend to cut small vertex sets / consider only

Ratio Cut : Minimize $\frac{cut(C_1, C_2)}{|C_1|} + \frac{cut(C_2, C_1)}{|C_2|}$

Normalized Cut : Minimize $\frac{cut(C_1, C_2)}{vol(C_1)} + \frac{cut(C_1, C_2)}{vol(C_2)}$

volume of a set of nodes

$\mathrm{vol}(C_i) =$
$assoc(C_i, V) =$
$cut(C_i, V) =$
$\sum_{v_i \in C_i, v_j \in V} w(v_i, v_j) =$
$\sum_{v_i \in C_i} \deg(v_i) =$

a) $Cut(C_1, C_2) = 4$

b) $\frac{Cut(C_1, C_2)}{|C_1|} + \frac{Cut(C_2, C_1)}{|C_2|} = \frac{5}{3} + \frac{5}{4} = \frac{35}{12}$

c) $\frac{Cut(C_1, C_2)}{Vol(C_1)} + \frac{Cut(C_1, C_2)}{Vol(C_2)} = \frac{5}{12+5} + \frac{5}{28+5}$

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) $|C_1|\sqrt{\frac{|\overline{C_1}|}{|C_1|}} + |\overline{C_1}| - \sqrt{\frac{|C_1|}{|\overline{C_1}|}}$

$= \sqrt{|C_1||\overline{C_1}|} - \sqrt{|\overline{C_1}||C_1|} = 0$

b) $(f_{C_1,i})^2 = |C_1|\frac{|\overline{C_1}|}{|C_1|} + |\overline{C_1}|\frac{|C_1|}{|\overline{C_1}|}$

$= |V|$

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

$= \frac{1}{2}\Big[\sum_{u\in C_1, v\in C_1} W_{uv}\cancel{(f_{C_1,u} - f_{C_1,v})^2} + \sum_{u\in\overline{C_1}, v\in\overline{C_1}} \cancel{(f_{C_1,u} - f_{C_1,v})^2}$

$+ 2\overline{\sum_{u\in C_1, v\in\overline{C_1}}} W_{uv}(f_{C_1,u} - f_{C_1,v})^2\Big]$

$= \overline{\sum} W_{uv}(f_{C_1,u}^2 - 2 f_{C_1,u} f_{C_1,v} + f_{C_1,v}^2)$

$\left(\frac{|\overline{C_1}|}{|C_1|} + 2 + \frac{|C_1|}{|\overline{C_1}|}\right) \underbrace{\overline{\sum} W_{uv}}_{\text{cut}(C_1,\overline{C_1})}$

$|V|\left(\frac{(|\overline{C_1}| + |C_1|)}{|C_1||\overline{C_1}|}\right)$

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

CT?

b) How does the spectral embedding change? move closer,

c) How does this change affect the final clustering? V_9 may inter V_6 cluster

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