Machine Learning for Graphs and Sequential Data Exercise Sheet 11

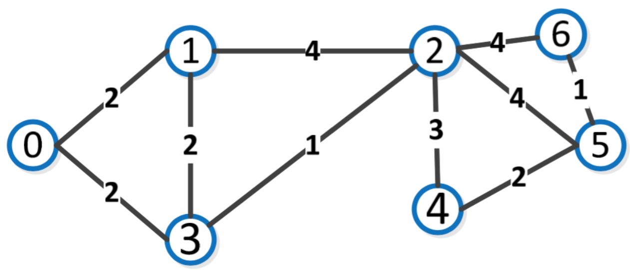
Graphs: Clustering

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

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

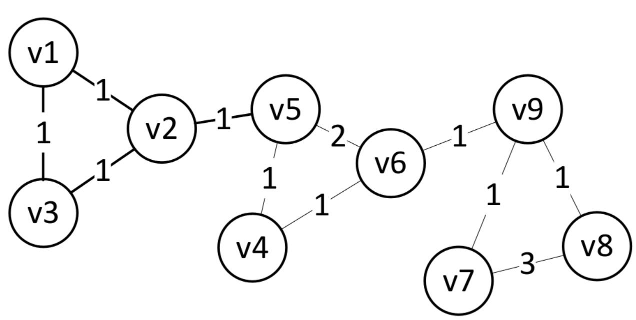
is defined as in the lecture. Prove the following three properties about $\boldsymbol{f}_{C_1}$.

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

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 4: Consider a PPM with intra-community edge probability p and inter-community edge probability q and a community assignment vector $\boldsymbol{z} \in \{1, -1\}^N$. Assume that the communities are balanced, i.e. $\sum_i z_i = 0$ and both communities have $\frac{N}{2}$ members. Show that the likelihood of an observed adjacency matrix $\boldsymbol{A} \in \{0, 1\}^{N \times N}$ is

$$\mathrm{p}(\boldsymbol{A} \mid \boldsymbol{z}) \propto \left(\frac{(1-p)q}{(1-q)p} \right)^{\mathrm{cut}(\boldsymbol{A};\boldsymbol{z})}$$

where

$$\mathrm{cut}(\boldsymbol{A}; \boldsymbol{z}) = \sum_{i<j} A_{ij} \cdot \mathbb{I}(z_i \neq z_j)$$

is the size of the cut in graph structure $\boldsymbol{A}$. Note that $\boldsymbol{A}$ is observed and thus fixed and we are interested in $\mathrm{p}(\boldsymbol{A} \mid \boldsymbol{z})$ as a function of $\boldsymbol{z}$.