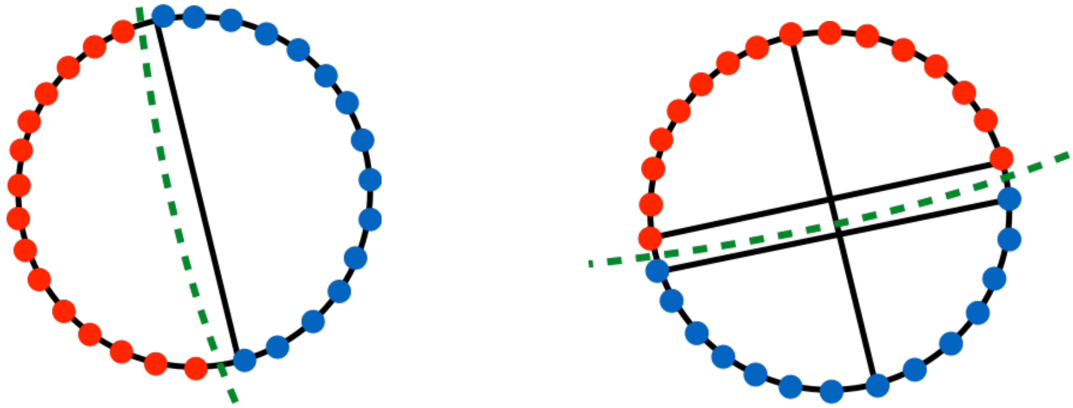
*Q1 (***Spectral Clustering Algorithm).**



The lefthand side in the above image indicates the initial graph and the righthand side indicates the new graph with two new edges added in. The green dotted line indicates the cuts obtained from spectral clustering.

Why it works: For the graph on the left in the figure, it is clear that the green line is the one that minimizes normalized cut. This is because we need to cut a minimum of two edges in the graph to even get a partition. The shown cut has a normalized cut value of:

$$\frac{N_{cut}}{N_1} + \frac{N_{cut}}{N_2} = \frac{2}{14} + \frac{2}{15}$$

Any other cut will only worsen the above quantity since we still will have a numerator of at least 2, but for the denominator, one term increases while other decreases by same quantity, and with that it is easy to verify that the shown cut is the one that minimizes the normalized cut. On the other hand once we add the two extra edges shown in the figure on the right, the original cut is no longer the optimal one. It cuts through 4 edges and has a value

$$\frac{4}{14} + \frac{4}{15}$$

However the cut indicated on the right only cuts 3 edges and has a normalized cut value of

$$\frac{3}{15} + \frac{3}{15} = \frac{2}{5}$$

Any other cut will only increase the normalized cut value. The key intuition is that we are looking for clusters that cut as few edges as possible and have a roughly equal number of edges within the two clusters.

*Q2 (***EM Algorithm for Mixture of Poisson).**

1. E-step for mixture model is given by (from lecture 14)

$$\begin{aligned} Q_t^{(i)}(k) &= P(c_t = k | X_t = x_t) \\ &\propto P(X_t = x_t | c_t = k) P(c_t = k) \\ &\propto (\lambda_k^{(i)})^{x_t} e^{-\lambda_k^{(i)}} \pi^{(i-1)}(k) \end{aligned}$$

   Hence we have that:

$$Q_t^{(i)}(k) = \frac{(\lambda_k^{(i)})^{x_t} e^{-\lambda_k^{(i)}} \pi^{(i-1)}(k)}{\sum_{c=1}^{K} (\lambda_c^{(i)})^{x_t} e^{-\lambda_c^{(i)}} \pi^{(i-1)}(c)}$$

2. For N-step, for mixture model we proved in class that

$$\pi^{(i)}(k) = \frac{1}{n} \sum_{t=1}^{n} Q_t^{(i)}(k)$$

3. Finally lets do M-step for λ_k. To do this, we need to optimize:

$$\begin{aligned} &\underset{\lambda_1,\ldots,\lambda_K}{\operatorname{argmax}} \sum_{t=1}^{n} \sum_{k=1}^{K} Q_t^{(i)}(k) \left(\log P(X_t = x_t | c_t = k) + \log \pi_k \right) \\ &= \underset{\lambda_1,\ldots,\lambda_K}{\operatorname{argmax}} \sum_{t=1}^{n} \sum_{k=1}^{K} Q_t^{(i)}(k) \left(\log P(X_t = x_t | c_t = k) \right) \\ &= \underset{\lambda_1,\ldots,\lambda_K}{\operatorname{argmax}} \sum_{t=1}^{n} \sum_{k=1}^{K} Q_t^{(i)}(k) \left(\log \left(\frac{\lambda_k^{x_t} e^{-\lambda_k}}{x_t!} \right) \right) \\ &= \underset{\lambda_1,\ldots,\lambda_K}{\operatorname{argmax}} \sum_{t=1}^{n} \sum_{k=1}^{K} Q_t^{(i)}(k) \left(x_t \log(\lambda_k) - \lambda_k \right) \\ &= \underset{\lambda_1,\ldots,\lambda_K}{\operatorname{argmax}} \sum_{k=1}^{K} \sum_{t=1}^{n} Q_t^{(i)}(k) \left(x_t \log(\lambda_k) - \lambda_k \right) \end{aligned}$$

   Now lets optimize for a particular λ_k by taking derivative and equating to 0. That is,

$$\sum_{t=1}^{n} Q_t^{(i)}(k) \left(\frac{x_t}{\lambda_k} - 1 \right) = 0$$

   Thus we conclude that,

$$\sum_{t=1}^{n} Q_t^{(i)}(k) \frac{x_t}{\lambda_k} = \sum_{t=1}^{n} Q_t^{(i)}(k)$$

   Or in other words,

$$\lambda_k = \frac{\sum_{t=1}^{n} Q_t^{(i)}(k) x_t}{\sum_{t=1}^{n} Q_t^{(i)}(k)}$$

   Thus for M-step we set each $\lambda_k^{(i)}$ as above.