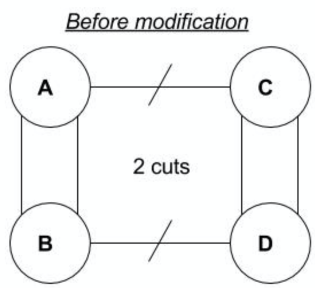
A2 Solution

1 Clustering Sensitivity



To generate the original data points, we first create four connected components with approximately equal nodes in each (8, 7, 8, 7). Let's call them A, B, C, and D. We then introduce **two edges** between (A, B), and (C, D) and **one edge** between (A, C) and (B,D). So now the graph is completely connected. We call this Initial Graph (AspectralI), as shown in the figure above. The actual graph can be seen in Figure 1.
Now, when we perform spectral clustering with K=2, the algorithm clusters A and B as one cluster (blue), and C and D as the other cluster (red) (shown in Figure 1) by cutting the single edge between (A, C) and (B, D). This is because spectral clustering tries to optimize the following criteria $CUT(c1, c2) * (\frac{1}{Edges(c1)} + \frac{1}{Edges(c2)})$.

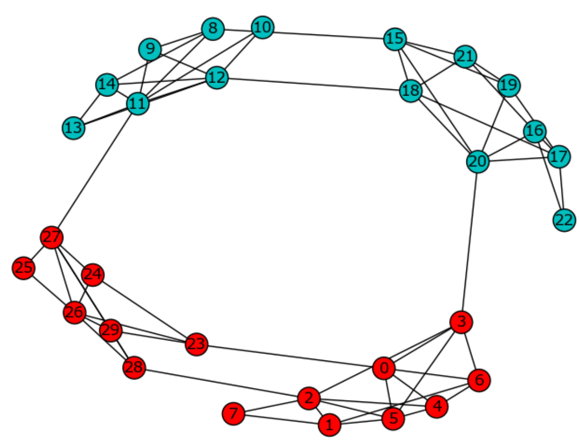
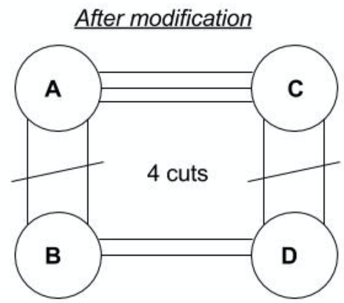


Figure 1: Initial Graph



To create the new graph, we then add **three new edges** between (A, C) and (B, D). So now we have two edges between (A,B) and (C,D), and three between (A,C) and two between (B,D) (shown in figure above). Now, when we run spectral clustering on this modified graph, the algorithm cuts the two edges between (A, B) and (C, D), total 4 edges are cut. The new clustering would then be A and C as one cluster (blue) and B and D (red) as the other cluster (shown in Figure 2). This means the cluster assignments for nodes in B and C have changed. So the labels of 15 out of the 30 nodes have changed, which is **50%**. The algorithm does not cut the two edges between B and D, because if it had the objective function as mentioned above would not have been minimized.

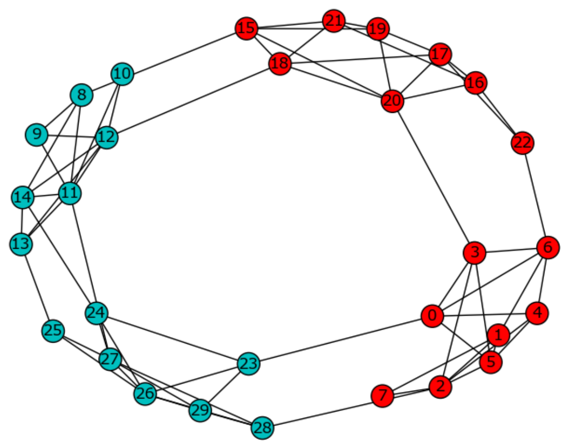


Figure 2: Modified Graph: Added three edges 13 < − > 25 & 6 < − > 22 & 14 < − > 24

The two figures below show the two-dimensional mapping of the data points (Initial Graph (Figure 3) and Modified graph (Figure 4)). We got these figures through the intermediate step in spectral clustering algorithm when we take eigen vector corresponding to two smallest eigen values. These figures provide more intuition.

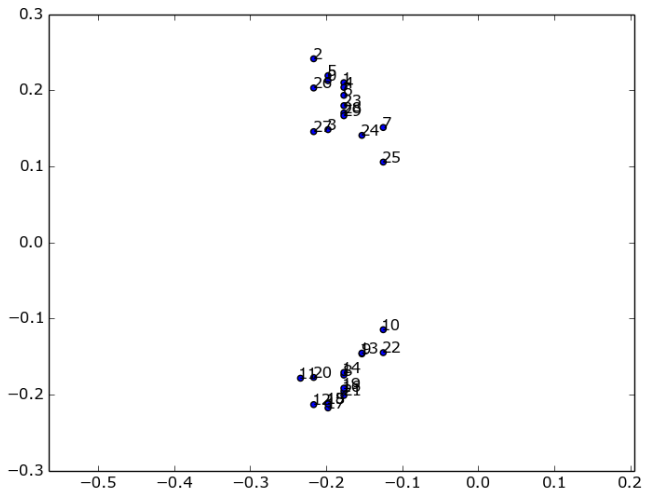


Figure 3: Initial Graph: Mapping in two dimensional space

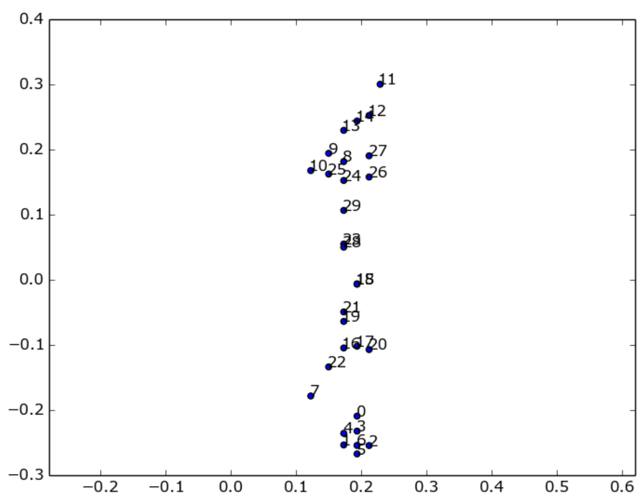


Figure 4: Modified Graph: Mapping in two dimensional space

2 EM for Mixture of Poisson distributions:

a) E-Step

For a given iteration i, for each data point $t \in$ [n],

$$Q_t^{(i)}(c_t) = P\Big(c_t \mid x_t \,,\, \theta^{(i-1)}\Big)$$

$$\propto \; p\Big(x_t \mid c_t \,,\, \theta^{(i-1)}\Big) \;\times\; P\Big(c_t \mid \theta^{(i-1)}\Big)$$

$$\propto \; p\Big(x_t \mid c_t, \theta^{(i-1)}\Big) \;\times\; P\Big(c_t \mid \theta^{(i-1)}\Big)$$

$$\frac{p\Big(x_t \mid c_t \,,\, \theta^{(i-1)}\Big) \cdot \pi^{(i-1)}[c_t]}{\sum_{c_t=1}^{K} p\Big(x_t \mid c_t \,,\, \theta^{(i-1)}\Big) \cdot \pi^{(i-1)}[c_t]} \tag{1}$$

In this case, we have a Poisson distribution $\Rightarrow [\theta] = [\lambda]$

$$p\Big(x_t \mid c_t \,,\, \lambda_{c_t}^{(i-1)}\Big) = p\Big(x_t \mid \lambda_{c_t}^{(i-1)}\Big)$$

$$= \frac{\lambda_{c_t}^{x_t} \; e^{-\lambda_{c_t}}}{x_t!} \tag{2}$$

Substituting (2) into (1),

$$Q_t^{(i)}(c_t) = \frac{\frac{\lambda_{c_t}^{x_t} \; exp(-\lambda_{c_t})}{x_t!} \cdot \pi^{(i-1)}[c_t]}{\sum_{c_t=1}^{K} \frac{\lambda_{c_t}^{x_t} \; exp(-\lambda_{c_t})}{x_t!} \cdot \pi^{(i-1)}[c_t]}$$

M-Step update for π and $\lambda_1, ..., \lambda_K$

Given $Q_1, ..., Q_n$,

$$\theta^{(i)} = \; \underset{\theta \epsilon \Theta}{\arg\max} \sum_{t=1}^{n} \; \sum_{k=1}^{K} Q_t^{(i)}(c_t = k) \; \log \; P\Big(x_t, c_t = k \mid \theta\Big)$$

$$= \; \underset{\theta \epsilon \Theta}{\arg\max} \; \sum_{t=1}^{n} \; \sum_{k=1}^{K} Q_t^{(i)}(c_t = k) \; \Bigg(\log \; P\Big(x_t | c_t = k, \theta\Big) + log \; P\Big(c_t = k \mid \theta\Big)\Bigg)$$

$$= \; \underset{\pi, \lambda_1, ..., \lambda_K}{\arg\max} \; \sum_{t=1}^{n} \; \sum_{k=1}^{K} Q_t^{(i)}(k) \; \Bigg(\; \log\Big(Poisson(x_t; \lambda_k)\Big) + \; \log \; \pi_k \Bigg)$$

$$= \; \underset{\pi, \lambda_1, ..., \lambda_K}{\arg\max} \; \sum_{t=1}^{n} \; \sum_{k=1}^{K} Q_t^{(i)}(k) \; \Bigg(\; \Bigg(\frac{\lambda_k^{x_t} e^{-\lambda_k}}{x_t!}\Bigg) + \; \log \; \pi_k \Bigg)$$

Let $L(\pi, \lambda_1, ..., \lambda_K) =$

$$\sum_{t=1}^{n} \sum_{k=1}^{K} Q_t^{(i)}(k) \left(\log \left(\frac{\lambda_k^{x_t} \, e^{-\lambda_k}}{x_t!} \right) + \log \pi_k \right)$$

b) M-Step update for π

Since we want to optimize over π subject to constraint:

$$\sum_{k=1}^{K} \pi_k = 1$$

So, we want to optimize $L(\pi, \lambda_1, ..., \lambda_k) \; + \; \lambda \left(1 - \sum_{k=1}^{K} \pi_k\right)$

As proved in class,

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

c) M-Step update for $\lambda_1, ..., \lambda_K$

We want to optimize over $\lambda_1, ..., \lambda_K$. There is no constraint on λ, as for any $\lambda > 0$, the Poisson distribution will sum up to 1.

So taking the derivative of $L(\pi, \lambda_1, ..., \lambda_K)$ w.r.t. λ_k and equating it to 0,

$$\frac{\delta L(\lambda_1, ..., \lambda_K)}{\delta \lambda_k} = 0$$

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

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

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

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