Homework 11 - Solution

1. (a) We have

$$\|y_l^t - \Phi(x_j)\|_{\mathcal{H}}^2 = \|\frac{1}{n_l^{t-1}} \sum_{i \in I_l^{t-1}} \Phi(x_i) - \Phi(x_j)\|_{\mathcal{H}}^2$$

$$= \|\frac{1}{n_l^{t-1}} \sum_{i \in I_l^{t-1}} (\Phi(x_i) - \Phi(x_j))\|_{\mathcal{H}}^2$$

$$= \frac{1}{(n_l^{t-1})^2} \sum_{i,i' \in I_l^{t-1}} \langle \Phi(x_i) - \Phi(x_j), \Phi(x_{i'}) - \Phi(x_j)\rangle_{\mathcal{H}}$$

$$= \frac{1}{(n_l^{t-1})^2} \sum_{i,i' \in I_l^{t-1}} (K(x_i, x_{i'}) + K(x_j, x_j) - K(x_i, x_j) - K(x_{i'}, x_j)).$$

The last equality is obtained by expanding the dot product and using the property $\langle \Phi(x), \Phi(x')\rangle_{\mathcal{H}} = K(x, x')$.

(b) For $t = 0$, the clusters I_l^0 are defined by $I_l^0 = \{l\}$ by definition: we can compute them. By induction, assume that we can compute the clusters I_l^{t-1}. Let us show that we can compute the clusters I_l^t. Using the formula proven in the last question, we can compute the distance between each observation $\Phi(x_j)$ and y_t^l while only knowing the set I_l^{t-1} (that we have access to by induction). Therefore, we can know for each j what is the centroid y_t^l that is the closest to $\Phi(x_j)$. We then compute the sets I_l^t by computing the indexes j such that $\Phi(x_j)$ is the closest to y_l^t. This ends the induction step.

2. (a) If $q = 0$, then there are two connected components (equal to $\{1, \ldots, n\}$ and $\{n+1, \ldots, 2n\}$. As long as $q > 0$, there is a single connected component.

(b) For $q = 0$, there is a single connected component. According to the lecture notes, this implies that 0 has multipicity one as an eigenvalue. The vector $(1, \ldots, 1)$ is the corresponding eigenvector.

(c) For $q > 0$, there are two connected components. Therefore, the eigenvalue 0 has multiplicity 2. An orthonormal basis of the associated eigenspace is given by $(1, \ldots, 1, 0, \ldots, 0)$ and $(0, \ldots, 0, 1, \ldots, 1)$ (each vector having n ones and n zeroes).

(d) We need to show that $Lu = \lambda u$. The degree matrix of the graph is the matrix $1/((q+1)n)\mathrm{Id}_n$. Therefore, $L = \mathrm{Id}_n - W/((q+1)n)$. This implies that having $Lu = \lambda u$ is equivalent to having

$$Wu = n(q+1)(1-\lambda)u.$$

Let us show this equality. Let $e \in \mathbb{R}^n$ be the vector with all ones, so that $u = (e, -e)$. We have

$$Wu = \begin{pmatrix} E_n & qE_n \\ qE_n & E_n \end{pmatrix} \begin{pmatrix} e \\ -e \end{pmatrix} = \begin{pmatrix} (1-q)E_n e \\ (q-1)E_n e \end{pmatrix} = n \begin{pmatrix} (1-q)e \\ (q-1)e \end{pmatrix}. \tag{1}$$

Also,

$$n(q+1)(1-\lambda)u = n \begin{pmatrix} (q+1)(1-\lambda)e \\ -(q+1)(1-\lambda)e \end{pmatrix} \tag{2}$$

One can check that for $\lambda = 2q/(q+1)$, it holds that $(q+1)(1-\lambda) = 1-q$ and $-(q+1)(1-\lambda) = q-1$. This concludes the proof.

(e) For $q \ll 1$, the eigenvalue $2q/(q+1)$ is very close to 0. Even if the eigenvalue 0 has a multiplicity 1 (there is a single connected component), the existence of this very small eigenvalue reflects the fact that there are two clusters in this graph: indeed, the similarity between two $i \leq n$ and $j > n$ is small (equal to $q \ll 1$), whereas the similarity in a cluster is equal to 1.