Quiz 8 Solution

Q1

(1pt) Adjacency Matrix (a_{ij}=*1* if edge between node *i* and *j*)

	1	2	3	4	5	6
1	0	1	0	0	1	0
2	1	0	1	0	1	0
3	0	1	0	1	0	0
4	0	0	1	0	1	1
5	1	1	0	1	0	0
6	0	0	0	1	0	0

(1pt) Degree Matrix (d_{ij}=*degree of node i*)

	1	2	3	4	5	6
1	2					
2		3				
3			2			
4				3		
5					3	
6						1

(1pt) Laplacian Matrix = Degree Matrix - Adjacency Matrix

	1	2	3	4	5	6
1	2	-1	0	0	-1	0
2	-1	3	-1	0	-1	0
3	0	-1	2	-1	0	0
4	0	0	-1	3	-1	-1
5	-1	-1	0	-1	3	0
6	0	0	0	-1	0	1

Q2 (1pt) Spectral Clustering Algorithms

1) Pre-processing: Construct a **matrix representation** of the graph
2) Decomposition: Compute **eigenvalues** and **eigenvectors** of the matrix; Map each point to a **lower-dimensional representation** based on one or more eigenvectors
3) Grouping: **Assign points** to two or more clusters, based on the new representation

Q3 Clustering Algorithm

a) (1pt)Given a method of presenting document to some space(e.g. vector/set) with a notion of distance between documents. Group documents into some number of clusters that the similar documents are assigned to the same cluster.

b) (method 0.5pt each; distance measure & explanation 0.5pt each)

- Set of words. Jaccard distance = 1 - $|X \cap Y| / |X \cup Y|$
- Vector in space of selected words. Cosine distance = 1 - $X \cdot Y/|X|*|Y|$
- Point in space of selected words. Euclidean distance:

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_i - q_i)^2 + \cdots + (p_n - q_n)^2}.$$

(Hamming Distance/ Edit Distance are also correct for vector)

Q4 (key operation 0.5pt each, implementation detail 0.5pt each)

1) Represent a cluster of many points. Centroid when measuring distance of clusters using Euclidean distance.
2) Determine the “nearness” of cluster. Measuring cluster distances by distances of centroids.
3) Stop combining clusters when cluster number reaching the intuitive number of clusters or best combination(keep merging leads to inadequate).