Kajol Taresh Shah
A20496724
Spring 2022

CS579 Online Social Network
Assignment 2.

Q.1.
a)

$A_{ij} = \begin{cases} 1, & \text{if there is an edge between } v_i \,\&\, v_j \\ 0 & \text{otherwise} \end{cases}$

	V1	V2	V3	V4	V5
V1	0	1	0	0	1
V2	1	0	1	0	1
V3	0	1	0	1	1
V4	0	0	1	0	1
V5	1	~~0~~ 1	1	1	0

Q1 b)

$$\lambda C_e = A^T C_e.$$

where λ is the eigenvalue matrix, A^T is adjacency matrix transpose & C_e eigen-vector
$A = A^T$ as this is an undirected graph. So,

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

The eigen values (λ) which we get are :-
-1.61803, 0.61803, -0.46259, -1.47283, 2.93543

Katz Centrality is given by

$$C_{katz} = \beta(I - \alpha A^T)^{-1} \cdot 1.$$

Here, $\alpha = 0.3$ & $\beta = 0.2$ is given.

$$C_{katz} = 0.2(I - 0.3 A^T)^{-1} \cdot 1$$

$$\cancel{C_k} (I - 0.3 A^T) = \begin{bmatrix} 1 & -0.3 & 0 & 0 & -0.3 \\ -0.3 & 1 & -0.3 & 0 & -0.3 \\ 0 & -0.3 & 1 & -0.3 & -0.3 \\ 0 & 0 & -0.3 & 1 & -0.3 \\ -0.3 & -0.3 & -0.3 & -0.3 & 1 \end{bmatrix}$$

$\det(I - 0.3A^T) \neq 0$ ($\det = 0.23716$)

So, inverse of $(I - 0.3A^T)$ exists

$$\therefore (I - 0.3A^T)^{-1} = \begin{bmatrix} 1.86 & 1.34 & 1.10 & 0.78 & 1.53 \\ 1.34 & 2.50 & 1.67 & 1.10 & 1.98 \\ 1.10 & 1.67 & 2.50 & 1.34 & 1.98 \\ 0.78 & 1.10 & 1.34 & 1.86 & 1.53 \\ 1.53 & 1.98 & 1.98 & 1.53 & 3.11 \end{bmatrix}$$

$$0.2(I - 0.3A^T)^{-1} = \begin{bmatrix} 0.372 & 0.268 & 0.22 & 0.156 & 0.306 \\ 0.268 & 0.5 & 0.334 & 0.22 & 0.396 \\ 0.22 & 0.334 & 0.5 & 0.268 & 0.396 \\ 0.156 & 0.22 & 0.268 & 0.372 & 0.306 \\ 0.306 & 0.396 & 0.396 & 0.306 & 0.622 \end{bmatrix}$$

Multiplying the above matrix with the vector of 1s, we get

$$0.2(I - 0.3A^T)^{-1} \cdot 1 = \begin{bmatrix} 0.372 & 0.268 & 0.22 & 0.156 & 0.306 \\ 0.268 & 0.5 & 0.334 & 0.22 & 0.396 \\ 0.22 & 0.334 & 0.5 & 0.268 & 0.396 \\ 0.156 & 0.22 & 0.268 & 0.372 & 0.306 \\ 0.306 & 0.396 & 0.396 & 0.306 & 0.622 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

C Katz.

$$\therefore C_{Katz} = \begin{bmatrix} 1.322 \\ 1.718 \\ 1.718 \\ 1.322 \\ 2.026 \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \longrightarrow \text{most important node.} \end{matrix}$$

Based on Katz centrality we rank the nodes as follow

$$\begin{bmatrix} 1.322 \\ 1.718 \\ 1.718 \\ 1.322 \\ 2.026 \end{bmatrix} \Rightarrow \begin{bmatrix} 3 \\ 2 \\ 2 \\ 3 \\ 1 \end{bmatrix} \text{ OR } \begin{bmatrix} 4 \\ 2 \\ 2 \\ 4 \\ 1 \end{bmatrix}$$

(either of the ways).

	Degree Centrality	Katz Centrality	Ranks (Katz)
V1	0.5	1.322	3
V2	0.75	1.718	2
V3	0.75	1.718	2
V4	0.5	1.322	3
V5	1	2.026	1

Degree centrality normalized by max degree :- (max degree = 4).

$V_1 = 2/4 = 0.5$
$V_2 = 3/4 = 0.75$
$V_3 = 3/4 = 0.75$
$V_4 = 2/4 = 0.5$
$V_5 = 4/4 = 1.$

Q1.c) The eigen values which we got are -1.61803, 0.61803, -0.46259, -1.47283, 2.93543. The highest eigen value is 2.93543. The value of α should be less than $1/\lambda$ i.e. $1/2.93543$.

$\alpha = 0.3 < \frac{1}{\lambda} = 0.34.$

In practice, $\alpha < 1/\lambda$ is selected so that centralities are computed correctly

Q1. d) If we set $\alpha = 0$, the eigenvector centrality will be removed and all the nodes in the graph will get the same centrality value which is 'β'.

Q1. e) Global clustering coefficient of the graph

$$C = \frac{\text{Number of Triangles} \times 3}{\text{Number of Connected Triples of nodes}}$$

Number of closed triples / triangles = 3

Number of connected triples of nodes :–
(open)

V_5 : $V_1 V_5 V_3$, $V_1 V_5 V_4$, $V_2 V_5 V_4$

V_2 : $V_1 V_2 V_3$

V_3 : $V_2 V_3 V_4$

$$\therefore C = \cancel{\frac{3\times3}{3\times3+4}} = \cancel{\frac{6}{10}} = \cancel{\frac{3}{5}} = \cancel{\frac{9}{13}}$$

$$\therefore C = \frac{3\times3}{3\times3+5} = \frac{9}{14} = 0.64$$

Q1. f) Cosine similarity :-

$$\sigma_{Cosine}(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{\sqrt{|N(v_i)||N(v_j)|}}$$

$N(V_2)$:- V_1, V_3, V_5
$N(V_5)$:- V_1, V_4, V_3, V_2

$$\sigma_{Cosine}(V_2, V_5) = \frac{|(V_1, V_3, V_5) \cap (V_1, V_4, V_3, V_2)|}{\sqrt{|(V_1, V_3, V_5)||(V_1, V_4, V_3, V_2)|}}$$

$$= \frac{2}{\sqrt{3 \times 4}}$$

$$= \frac{2}{\sqrt{12}} = \frac{2}{2\sqrt{3}} = \frac{1}{\sqrt{3}} = 0.57.$$

Q2. a) Random graphs follow the basic assumption that Edges (i.e friendship) between nodes (i.e individuals) are formed randomly. But in reality, friendships in real world networks are far from random. By assuming random friendships, we simplify the process of friendship formation in real world networks hoping that the random friendships ultimately create networks that exhibit common characteristics observed in real world networks. ~~But~~ Random graphs perform well in modeling the average path lengths however, when considering the transitivity, the random graph model drastically underestimates the clustering coefficient.

So, random graphs are incapable of modeling real world graphs ~~to~~ and to tackle this issue, small world models are considered.

Q2.b) Local clustering coefficient is defined as

$$C(V_i) = \frac{(\text{Number of pairs of Neighbours of } V_i \text{ that are connected})}{\text{Number of pairs of neighbours of } V_i} \quad \text{— (1)}$$

Number of pairs of neighbours of V_i

$$= \frac{c(c-1)}{2} \quad \text{— (2)}$$

In a regular lattice, number of connections between neighbours is given by $\frac{3}{8} c(c-2)$ — (3)

Putting (3) & (2) in (1), we get

$$C(V_i) = \frac{(3/8)\, c(c-2)}{\frac{c(c-1)}{2}}$$

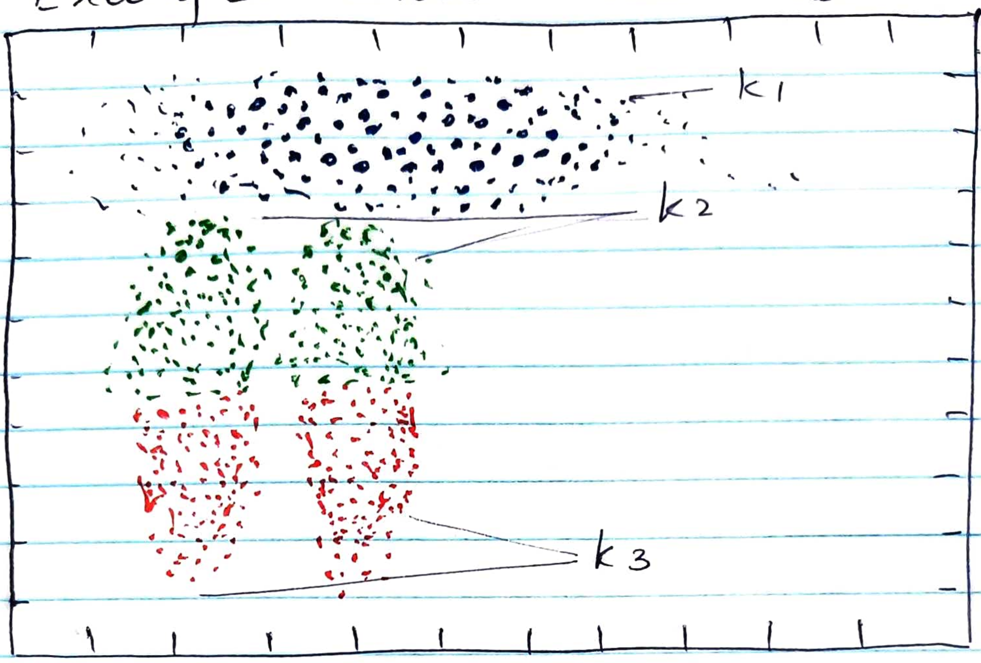
Rearranging,

$$C(V_i) = \frac{3 \times 2 \times c(c-2)}{8\, c(c-1)} = \frac{3(c-2)}{4(c-1)}$$

Hence, it is proved that local clustering coefficient for any node is $\frac{3(c-2)}{4(c-1)}$ where c is the average degree.

Q3. a) The usual shape of clusters generated by k-means is spherical or ball shaped. Sphere has same radius in each dimension. The k-means clustering uses Euclidean distance to cluster the data points. The Euclidean distance entails that the average of the coordinates of data points in a cluster is the centroid of that cluster. Clusters are modeled only by the position of their centroids and assumes all clusters have same radius and must occupy the same volume in data space. The number k of groupings in the data is fixed and known which is rarely the case in practice. Thus, k-means is quite inflexible and degrades badly when the assumptions are violated. So, when clusters are in different shapes such as elliptical clusters k-means doesn't work well.

Q 3.b) Example where k-means doesn't classify correctly



Here, the k-means algorithm is unable to classify data points correctly as k-means assumes spherical shape of clusters and so, it fails when clusters are in different shapes like elliptical. K-means uses centroid & Euclidean distance for clustering and so, it doesn't work here. Here, we can see that clusters k2 & K3 were not formed correctly as their shape was not spherical