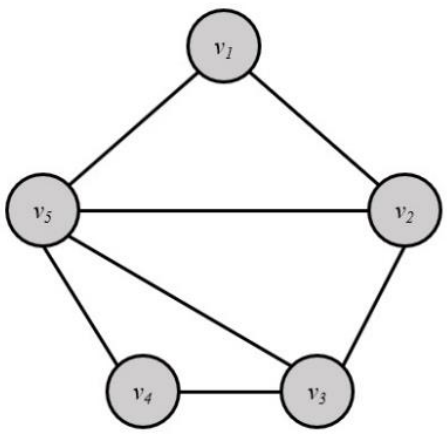
<u>**HW2**</u>

**1.** [Network Measures] Based on the following network answer the questions,

**(a)** Fill the adjacency matrix.




**Answer:**

| | 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 | 1 | 1 | 1 | 0 |

Here, the 1 means direct edge exists

The 0 means no direct edge exists.

**(b)** Calculate the "Degree Centrality" (normalized by the maximum degree) values and "Katz Centrality" values with α = 0.3 and β = 0.2, and rank the nodes based on Katz Centrality (you can use Matlab or other mathematical software to calculate the eigenvalues).

**Answer:**

<u>**Degree Centrality**</u> (Normalized by max degree]

V1 = $\frac{2}{4} = \frac{1}{2}$

V2 = $\frac{3}{4} = \frac{3}{4}$

V3= $\frac{3}{4} = \frac{3}{4}$

V4 = $\frac{2}{4} = \frac{1}{2}$

V5 = $\frac{4}{4} = 1$

The node "v5" has the Max Degree = 4.

**Katz Centrality**

Given, α =0.3, β = 0.2

| | | Rank |
|---|---|---|
| V1 | 1.32 | 4 |
| V2 | 1.72 | 2 |
| V3 | 1.72 | 2 |
| V4 | 1.32 | 4 |
| V5 | 2.01 | 1 |

Undirected graph: $A = A^T$

The eigen values are:

$$\mathbf{C}_{Katz} = \beta(\mathbf{I} - \alpha A^T)^{-1} \cdot \mathbf{1}$$

Eigen values are: (2.93,0.61,-0.46,-1.61,-1.47)

Here we will take the highest eigen value= 2.93

Therefore, the given value of α =0.3 and β = 0.2, where α =0.3 < $\frac{1}{2.935}$, now the condition satisfied.

$$\mathbf{C}_{Katz} = \beta(\mathbf{I} - \alpha A^T)^{-1} \cdot \mathbf{1}$$

= $0.2(I - 0.3\, A^T)^{-1}$ here "I" is identity matrix, $A=A^T$ is adjacency matrix.

$$= 0.2 * \left( \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} - 0.3 * \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix} \right)^{-1}$$

= [ 1.326, 1.725, 1.725,1.326,2.031]

| | Degree Centrality | Katz Centrality | Rank |
|---|---|---|---|
| V1 | ½ | 1.3265 | 4 |
| V2 | ¾ | 1.7245 | 2 |
| V3 | ¾ | 1.7245 | 2 |
| V4 | ½ | 1.3265 | 4 |
| V5 | 4/4 | 2.0306 | 1 |

**(c)** Is the above alpha value a good choice for Katz centrality? Why?

**Answer:**

Yes, the , α is good choice and it is accepted.

$$\alpha < \frac{1}{\lambda}$$

This condition should be True and it is True for the above values.

Given, α =0.3,β = 0.2

$$\alpha = 0.3 < \frac{1}{2.935}$$

So wew can say that this is a good choice.

**(d)** Discuss what would happen if we set α = 0?

**Answer**:

If α = 0, ketz centrality equal to β for all nodes. The eigen vector centrality part gets removed by making α = 0.And then all the nodes will have the some katz centrality which will be equal to β.

**(e)** Calculate the global clustering coefficient of the graph.

**Answer:**

Global clustering co-efficient of the graph is given by:

$$C = \frac{(Number\ of\ Triangles)*3}{No.Of\ connected\ nodes}$$

$$= \frac{3*3}{(3*3)+(5)}$$

$$= \frac{9}{14}$$

$$c = \frac{9}{14}$$

**(f)** Compute the similarity between nodes v2 and v5 using cosine similarity

**Answer:**

Cosine similarity between v2 and v5 is,

$$\Sigma_{\text{cosine}}(vi,vj) = \frac{|N(vi) \wedge N(Vj)}{\sqrt{|N(Vi)||N(Vj)}}$$

$$= \frac{|\{v1,v3,v5\} \wedge \{v1,v2,v3,v4\}}{\sqrt{|\{v1,v3,v5\}|\ |\{v1,v2,v3,v4\}|}}$$

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

$$= \frac{\sqrt{3}}{3} = 0.577$$

**2.** [Network Models]

**(a)** Why are random graphs incapable of modeling real-world graphs?

**Answer:**

Random graphs are regularly used as a simple mathematical model to examine graph features. However, for a number of reasons they may be incapable of fully representing real-world graphs. Random graphs are incapable of modeling real world graphs because real world graphs contain few connections and the nodes are not well connected. Real-world graphs may have certain characteristics that are not present in random graphs.

Random graphs have a short average path length and a small clustering co-efficient, but real-world graphs have a longer average path length, a larger clustering co-efficient, and a power law degree distribution. Additionally, the degree distribution of random graphs does not follow a power law.

**(b)** Show that in a regular lattice for small-world model, local clustering coefficient for any node is $\frac{3(c-2)}{4(c-1)}$ where c is the average degree. Hint: See problem 5 in the textbook.

**Answer:**

No.of triangles for a given node = No.of distinct ways to choose 2 forword vertices from $\frac{c}{2}$ possibilities

$$= \frac{\textit{Pairs of neighbours of vi (connected)}}{\text{Pairs of neighbours}}$$

$$= \frac{\left(\frac{c/2}{2}\right) x^3}{(c/2)}$$

$$= \frac{\left(\frac{c}{2}\right)\left(\frac{c}{2}-1\right) * \frac{3}{2}}{c(c-1)*2}$$

$$= \left(\frac{3}{4}\right)\frac{(c-2)}{(c-1)}$$

Therefore Proved.

**3.** [Unsupervised Learning]

**(a)** What is the usual shape of clusters generated by k-means? Why is this the case? Justify your answer by referencing the algorithm

**Answer:**

The clusters generated by the k-means spherical or a circular shape because it uses Euclidean distance metric and the objective is to minimizing the sum of squared distances between data points and centroids.. The K-mans algorithm minimizes the distance between the center of the cluster and the points that belong to the cluster. The shape of the clusters generated by k-means depends on the distance metric used to measure the distance between data points and centroids.

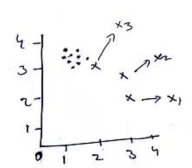
The clustures are spheres most of the time, the radius of the cluster is distance between the centroid and the point farthest from the centroid in the cluster. The k-means compute  the squared distance between the data point and the centroid, to enable clusters to obtain global maxima. If the distance between the  data point and the centroid is maximum compared to other clusters centroids then it is assigned to the particular cluster it belongs .

**(b)** Give (or draw) an example of the case where k-means is unable to correctly classify data instances due to the pattern of these instances. Hint: use your answer from part (a).

**Answer:**

The K-means cannot classify the data instances correctly when the data clusters result in complecated geometric shapes, anything non spherical(gobuler). K-means may also fail when there are outliers and clusters with differing densities.

Example:




The K-means consider the Point x2 close to x1 than to x3 since they have non spherical shape in the following example.

The K-means algorithm will not let the data points to share the same cluster if they are far from each other, even if they belong to same cluster.

So, due to this problem k-means will put X1 and X2 will be in the same cluster but data Point X3 will be in part of different cluster which it doesn't belong.