Name:	SOLUTION
IIT ID Number:	× × × × × × × × ××

CS 579: Online Social Network Analysis

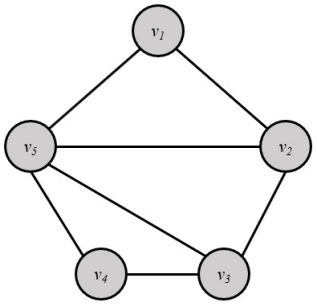
Homework II - Network Measures, Network Models and Clustering

Prof. Kai Shu
Due at 2022 Feb. 21, 11:59 PM

This is an ***individual*** homework assignment. Please submit a digital copy of this homework to **Blackboard**. For your solutions, even when not explicitly asked you are supposed to concisely justify your answers.

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

   (a) Fill the adjacency matrix.



	v_1	v_2	v_3	v_4	v_5
v_1	0	1	0	0	1
v_2	1	0	1	0	1
v_3	0	1	0	1	1
v_4	0	0	1	0	1
v_5	1	1	1	1	0

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

	Degree Centrality	**Katz Centrality**	**Ranks (Katz)**
v_1	$\frac{2}{4}$	1.3265	4
v_2	$\frac{3}{4}$	1.7245	2
v_3	$\frac{3}{4}$	1.7245	2
v_4	$\frac{2}{4}$	1.3265	4
v_5	1	2.0306	1

Name:	SOLUTION
IIT ID Number:	× × × × × × × × ××

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

Yes. Since in practice, $\alpha < \frac{1}{\lambda}$ is selected so that centralities are computed correctly.

(d) Discuss what would happen if we set $\alpha = 0$?

When $\alpha = 0$, the eigenvector centrality part is removed, and all nodes get the same centrality value β.

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

$$C = \frac{\text{(Number of Triangles)}\times 3}{\text{Number of Connected Triples of Nodes}} = \frac{3\times 3}{3\times 3+\underbrace{5}_{\{v_1,v_5,v_3\},\{v_1,v_5,v_4\},\{v_1,v_2,v_3\},\{v_2,v_5,v_4\}\{v_2,v_3,v_4\}}}$$

$C = \approx 0.64$

(f) Compute the similarity between nodes v_2 and v_5 using cosine similarity.

$$\sigma_{cosine}(v_i, v_j) = \frac{|N(v_i)\cap N(v_j)|}{\sqrt{|N(v_i)||N(v_j)|}} = \frac{|\{v_1,v_3,v_5\}\cap\{v_1,v_2,v_3,v_4\}|}{\sqrt{|\{v_1,v_3,v_5\}|\cdot|\{v_1,v_2,v_3,v_4\}|}} = \frac{2}{\sqrt{3\times 4}} = \frac{\sqrt{3}}{3}$$

2. **[Network Models]**

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

Real-world graphs have low average path length, high clustering coeffcient and power-law degree distribution while random graphs have low average path length and small clustering coefficient. Moreover, their degree distribution is not power-law at all.

(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.

Refer to the attached slides.

Name:	SOLUTION
IIT ID Number:	× × × × × × × × ××

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.

K-means clusters are Spherical. The algorithm minimizes the distance between the center of the cluster and the points belong to that cluster (in other words, radial distance of the points of the cluster to center of that cluster is similar).
From a probabilistic view, K-means assumes that the prior probability of the clusters and the variance of features are the same.

(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).

If clusters that need to be extracted are nonspherical (globular), k-means has problems detecting them. Moreover, K-means might fail in situations with outliers and clusters with different densities.