Name: ____________________ Roll #: ____________________ Section: ______________

National University of Computer and Emerging Sciences, Lahore Campus

	Course:	**Social Network Analysis**	**Course Code:**	**DS5115**
	Program:	**MS(Data Science)**	**Semester:**	**Spring 2019**
	Duration:	**180 Minutes**	**Total Marks:**	**100**
	Paper Date:	**29-May-19**	**Weight**	**50%**
	Section:	**ALL**	**Page(s):**	**10**
	Exam:	**Final Exam**		

Instruction/Notes: Attempt the examination on the question paper and write concise answers. You can use extra sheet for rough work. Do not attach extra sheets used for rough with the question paper. Don't fill the table titled Questions/Marks.

Question	1	2	3	4	5	6	7	Total
Marks	/ 15	/ 15	/ 20	/10	/20	/10	/10	/ 100

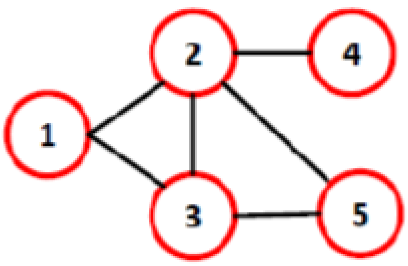
Q1) We have two simple, undirected graphs G1(V1, E1) and G2(V2, E2) with the same number of nodes |V1|=|V2| and the same number of edges |E1|=|E2|. G1 is a Watt-Strogatz random graph, while G2 has a power law degree distribution. Consider a virus v that starts at a random node and spreads according to the SIR model.

On which graph is there more likely to be an epidemic (defined as infecting at least 40% of the network)? Explain why in 1-2 sentences. **[5 Marks]**

In cases that an epidemic does take off, which graph will have a higher final percentage of nodes infected (on average)? Explain why in 1-2 sentences. **[5 Marks]**

Now the virus begins at the highest degree node in G 1 and G 2, respectively. On which graph is there more likely to be an epidemic? Explain why in 1-2 sentences. [**5 Marks**]

Q2) For the graph given below, Compute the node centrality and determine the ranking of the vertices according to each of the following centrality measures. Show your complete working.

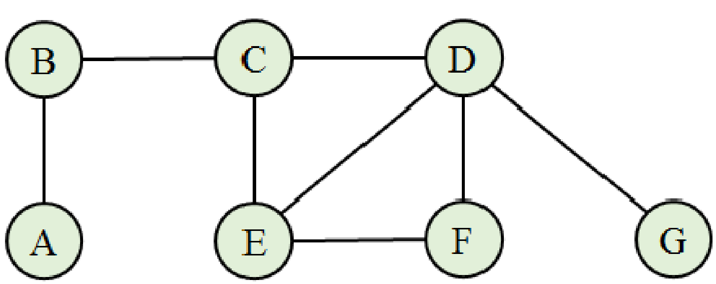


Harmonic Centrality [5 Marks]

Kat'z Index use β=1 and α=0.15. Show rerults uptil k=3 [5 Marks]

Betweeness Centrality [5 Marks]

Q3) Consider the following netwrok.



Compute the following local features for each node. **[5 Marks]**

1. The degree of v, i.e., deg(v)
2. The number of edges in the egonet of v, where egonet of v is defined as a subgraph whose nodes are v, and its neighbors and edges are induced from the whole graph
3. The number of edges that connects the egonet of v and the rest of the graph, i.e., the number of edges that enters or leaves the egonet of v.

For any pair of nodes u and v, we can use cosine similarity to measure how similar two nodes are according to their feature vectors x and y:

$$\text{Sim}(x, y) = \frac{x \cdot y}{||x||_2 \cdot ||y||_2} = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2} \cdot \sqrt{\sum_i y_i^2}};$$

Also, when $||x||_2 = 0$ or $||y||_2 = 0$, Sim(x; y) = 0. For node with ID E report the top 3 nodes that are most similar to node E using cosine similarity. Also mention their cosine similarity with node E. [**5 Marks**]

Next, we recursively generate some more features. We use mean and sum as aggregation functions. Our initial feature vector for node u is: $V_u \in \mathbb{R}^3$. In the first iteration, we concatenate the mean of all u's neighbors' feature vectors to V_u, and do the same for sum i.e.:

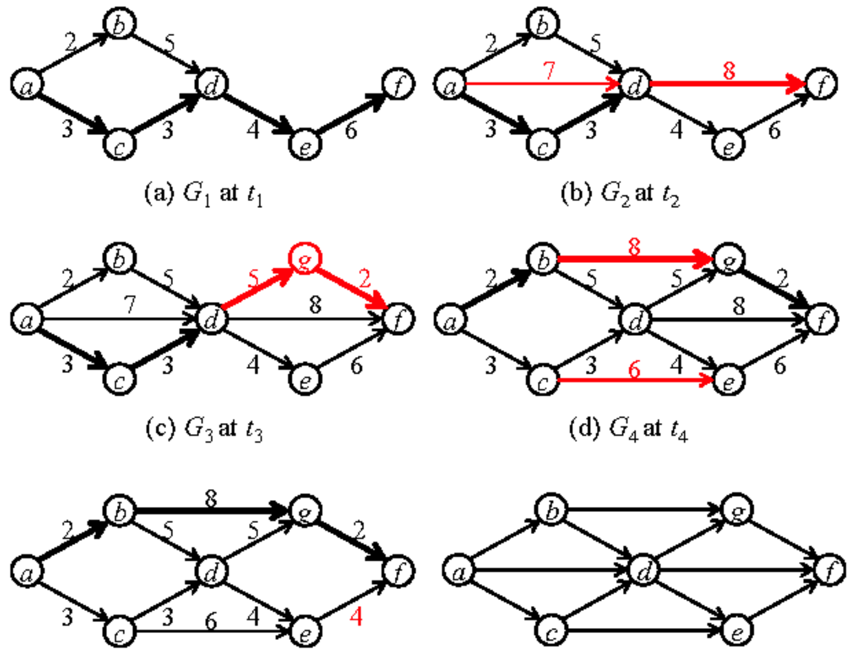
$$\tilde{V}_u^{(1)} = [\tilde{V}_u; \frac{1}{|N(u)|} \sum_{v \in N(u)} \tilde{V}_v; \sum_{v \in N(u)} \tilde{V}_v] \in \mathbb{R}^9,$$

Where N(u) is the set of u's neighbor in the graph. If N(u) is empty then set mean and sum value to 0.

Compute the feature vector $V_u^{(1)}$ for each node. **[5 Marks]**

Using this new feature vector, Report the top 3 nodes that are most similar to node E. Also mention their cosine similarity with node E. **[5 Marks]**

Q4) Consider the temporal graph below for indiviual time stamps.



(a) G_1 at t_1 (b) G_2 at t_2

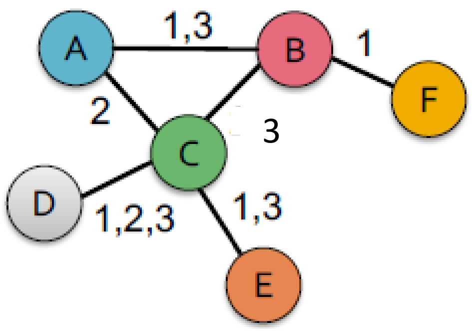
(c) G_3 at t_3 (d) G_4 at t_4

(e) G_5 at t_5 (f) super graph TEG

What is shortest path between node d and f at each time stamp i.e at t_1, t_2, t_3, t_4 and t_5. Also mention the distance between them. [**5 Marks**]

Name: ____________________ Roll #: ____________________ Section: ______________

Consider the following temporal network. Compute the temporal betweenness of node C til time t=2. Show yur working. [**5 Marks**]



Q5) Consider the following cascade model. each individual ***i*** has a threshold t_i that determines his/her behavior in the following way. If there are at least t_i individuals that are active, then he/she will be active, otherwise he/she stays inactive. It is assumed that each indiviual knows about the behavior of every one in the population (active/inactive). Nodes/indiviuals with $t_i=0$ are early adopters. The ones with $t_i > 0$ are late adopters.

In this problem we want to explore the impact of theshold distribution to find the total number of active indiviuals in a population of size n. You are given a histogram of thresholds $N = \{N_0, N_1, \ldots, N_{n-1}\}$ where N_i represents the number of people that have threshold i.

For the given Histogram, find the condition that checks whether the user with threshold *h* will be active or not. **[5 Marks]**

Write an algorithm that given a histogram N, computes the total number of active users of the population. Give high level pseudo code **[10 Marks]**

For the histogram given below what would be the output of your algorithm. **[5 Marks]**

Threshold	**# of people**
0	2
1	1
2	0
3	1
4	1
5	0
6	2
7	1
8	1
9	1

Q6) Suppose web pages are represented by numbers and you have already computed the personalized PageRank vectors for the following users:
U1, whose interests are represented by the teleport set {1, 2, 3}
U2, whose interests are represented by the teleport set {3, 4, 5}
U3, whose interests are represented by the teleport set {1, 4, 5}
U4, whose interests are represented by the teleport set {1}

Assume that the weights for each node in a teleport set are uniform. Without looking at the graph, can you compute the personalized PageRank vectors for the user U5 whose interests are represented by teleport set {2}? If so, how? If not, why not? Assume a fixed teleport Probability β. **[10 Marks]**

Q7) Recall that in Louvain method for community detection, when an isolated node i is added to a community C, the change in modularity is given by:

$$\Delta Q(i \rightarrow C) = \left[\frac{\sum_{in} \quad +k_{i,in}}{2m} - \left(\frac{\sum_{tot} \quad +k_i}{2m}\right)^2\right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m}\right)^2 - \left(\frac{k_i}{2m}\right)^2\right]$$

Where
$\sum_{in}$ = sum of link weights between nodes in C
$\sum_{tot}$ = sum of all link weights of nodes in C
$k_{i,in}$ = sum of weights of links from i to nodes in C
k_i = sum of all link weights of node i
Also need to derive $\Delta Q(D \rightarrow i)$ of taking node i out of community D

What would be the change in modularity when a node i is removed from the community D? **[10 Marks]**