Roll No.													

PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING

TEST 1

Sem & AY: Odd Sem 2019-20 | **Date**: 28.09.2019

Course Code: CSE 404 | **Time**: 9:30 to 10:30 AM

Course Name: SOCIAL NETWORK ANALYTICS | **Max Marks**: 40

Program & Sem: B.Tech & VII OE | **Weightage**: 20%

Instructions:

(i) Read the question properly and answer accordingly

(i) Question paper consists of 3 parts.

(ii) *Scientific and Non-programmable calculators are permitted.*

Part A [Memory Recall Questions]

Answer all the Questions. Each Question carries six marks. (3Qx6M=18M)

1. Define 'density' of a network with suitable example. (C.O.NO.1) [Knowledge]

2. Describe the steps to compute Eigen vector centrality of a node in a network graph. (C.O.NO.1) [Knowledge]

3. State the differences between walks, trails and paths in a network graph with proper illustration. (C.O.NO.1) [Knowledge]

Part B [Thought Provoking Questions]

Answer both the Questions. Each Question carries five marks. (2Qx5M=10M)

4. 'Closeness centrality measures global position of a node, whereas degree centrality captures such position locally' – review the statement. (C.O.NO.1) [Comprehension]

5. Eigenvector centrality overcomes one important drawback of other centrality measures-discuss. (CO.NO.1) [Comprehension]

Part C [Problem Solving Questions]

Answer both the Questions. Each Question carries six marks. **(2Qx6M=12M)**

6. Interpret the network data of the following directed graph (Figure 1) in the form of an edgelist and an adjacency matrix. (C.O.NO.1) [Application]

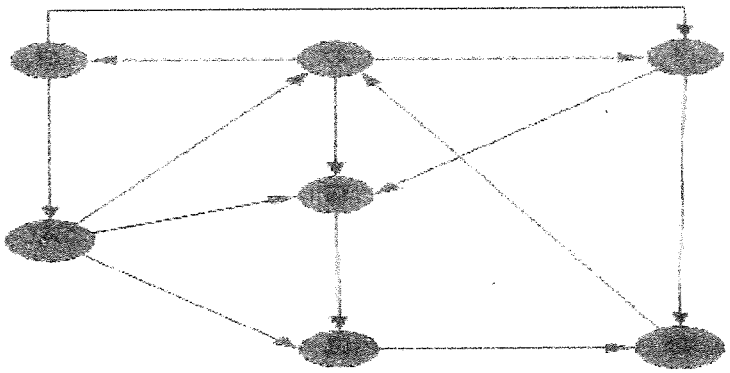


Figure 1

7. For the following network graph (Figure 2), compute Betweenness for node 2 and node 3. (C.O.1) [Application]

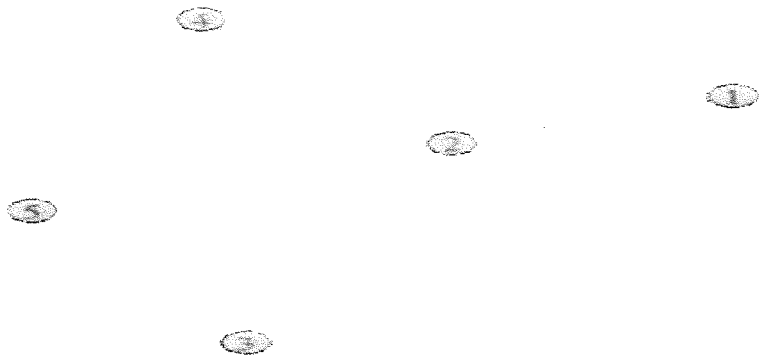


Figure 2

SCHOOL OF ENGINEERING

TEST – 1

Semester: VII
Course Code: CSE 404
Course Name: Social Network Analytics
Program & Sem: B.Tech & VII

Date: 28.09.2019
Time: 9:30 to 10:30 AM
Max Marks: 40
Weightage: 20%

Extract of question distribution [outcome wise & level wise]

Q.NO	C.O.NO	Unit/Module Number/Unit /Module Title	Memory recall type [Marks allotted] Bloom's Levels			Thought provoking type [Marks allotted] Bloom's Levels			Problem Solving type [Marks allotted]			Total Marks
			K	C	A	K	C	A	K	C	A	
1	CO1	Module 1	06	-	-							6
2	CO1	Module 1	06	-	-							6
3	CO1	Module 1	06	-	-							6
4	CO1	Module 1				-	05	-				5
5	CO1	Module 1				-	05	-				5
6	CO1	Module 1							-	-	06	6
7	CO1	Module 1							-	-	06	6
	Total Marks		18				10				12	40

K =Knowledge Level C = Comprehension Level, A = Application Level

Note: While setting all types of questions the general guideline is that about 60%

Of the questions must be such that even a below average students must be able to attempt, About 20% of the questions must be such that only above average students must be able to attempt and finally 20% of the questions must be such that only the bright students must be able to attempt.

[I hereby certify that All the questions are set as per the above guide lines. Mr. Tapas Guha]

Reviewers' Comments

Annexure- II: Format of Answer Scheme

SCHOOL OF ENGINEERING

SOLUTION

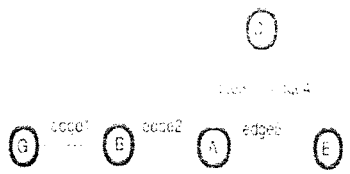
Semester: VII
Course Code: CSE 297
Course Name: Social Network Analytics
Program & Sem: B.Tech & VII

Date: 28.09.2019
Time: 9:30 to 10:30 AM
Max Marks: 40
Weightage: 20%

Part A (3Q x 6M = 18 Marks)

Q No	Solution	Scheme of Marking	Max. Time required for each Question
1	Basic description • number of nodes (n/V) • number of edges (m/E) • directed/undirected/bipartite/weighted (note: in simple graphs there are no multiple edges (ie. several undirected edges between the nodes) and self-loops (eg. edges from node A to A) $\text{density} \frac{\text{Number of edges present in the network}}{\text{Number of all possible edges in the network}} \quad \frac{m}{n(n-1)/2}$ For instance, for a network G with V(G)=35, E(G)=110, network density 110/(35(35-1)/2)=0.05 Density of 0.05 represents the ratio of links present in relation to all that could be (with the maximum density of 1. Density metric is highly dependent on the order and size of the network. Higher density indicates higher level of interconnectivity in the network (ie. many different ways as to how whatever travels through the network can be exchanged. Interconnectivity should be interpreted in relation to the phenomenon the network represents	3.5 marks for explanation + 2.5 marks for example.	6 mins

2	**Eigenvector centrality**
	• Suggested by Bonacich (1972)
	• Problem: what if the node has high degree but all of its connections are not well-connected to the rest of the network. Shouldn't centrality be defined by how well-connected your connections are?
	• Eigenvector centrality = A node is central to the extent that the node is connected to others who are central
	• Eigenvector centrality measure is based on walks
	  • Starts by assigning centrality score of 1 to all nodes (vi = 1 for all i)
	  • Recomputes scores of each node as weighted sum of centralities of all nodes in a node's neighborhood: vi = xijvj
	  • Normalizes v by dividing each value by the largest value
	  • Repeat steps 2 and 3 until values of v stop changing
		2 marks for basic discussion + 4 marks for computational steps.	6 mins

3	 • Walk (nodes and edges can be repeated): B -> A -> D -> A (what is passed through the network goes through the same node A twice and through edge 3 twice) • Trail (edges can not be repeated but nodes can): A -> D -> E -> A -> B (what is passed through the network does not pass the same edge twice but passes through the same node twice) • Path (no edges nodes can be repeated): B -> A -> E -> D (no edge or node is repeated, what is passed through the network goes through the unique sequence of nodes and edges)	**3 marks for difference + 3 marks for proper illustration.**	6 mins

Part B (2Q x 5M = 10 Marks)

Q No	Solution	Scheme of Marking	Max. Time required for each Question
4	**Degree** is a node-level metric describing the number of nodes adjacent to each of the nodes (or the number of links each node has). In directed networks, one **also** distinguishes in-degree (number of incoming ties) and out-degree (number of outgoing ties). A self-loop (i.e. retweeting one's own tweet in a communication network) contributes +2 to the degree as it counts +1 for both out-degree and for +1 in-degree. Thus it captures data local to that particular node. Farness is a sum of geodesic distances from each node to all the other nodes in the network, so the metric shows how far each node is from all the others. Closeness is inverse of farness, so it captures how close the node is **to all the others**, thus making it a global measure. Naturally, Closeness measures global position of a node, whereas degree captures such position locally.	**1 mark for each centrality + 3 for proper review.**	8 mins
5	The centrality of a node should be defined by how well-connected the connections of that node are, as in case of citation network or 'Friend of a Friend concept'. (have to explain a little). This is not taken into account by degree, betweenness	**5 for proper discussion.**	7 mins

and closeness centrality measures. Only eigenvector centrality pays heed to this concept thus overcoming this drawback of other centralities.

Part C (2Q x 6M = 12 Marks)

Q No	Solution	Scheme of Marking	Max. Time required for each Question
6	Given network graph is a directed one. so corresponding Edgelist:	2 marks for correct edgelist + 4 marks for correct adjacency matrix	10 mins

From	To
A	C
A	D
A	E
B	A
B	F
C	B
C	D
C	F
D	E
E	G
F	D
F	G
G	C

Corresponding **adjacency matrix**.

	A	B	C	D	E	F	G
A	0	0	1	1	1	0	0
B	1	0	0	0	0	1	0
C	0	1	0	1	0	1	0
D	0	0	0	0	1	0	0
E	0	0	0	0	0	0	1
F	0	0	0	1	0	0	1
G	0	0	1	0	0	0	0

3 for each correct computation | 12 mins

Betweenness for node 2:

- one geodesic from 1 to 4, and goes through 2 -> 1/1
- one geodesic from 1 to 3, and goes via 2 -> 1/1
- two paths from 1 to 5, and both through 2 -> 2/2
- two paths from 4 to 3, but only one via 2 -> 1/2

Betweenness for node 2:
sum(1/1+1/1+2/2+1/2)=3.5

Betweenness for node 5:

Sum (0/1 (*for path from 1 to 4*) +0/1 (*for path from 3 to 1*) + 1/2 (*for two paths from 3 to 4*) + 0 (*for all other paths that also do not go through node 5, and hence this fraction would equal to 0*) = 0.5.

Roll No.													

PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING

TEST – 2

Sem & AY: Odd Sem 2019-20 | **Date**: 20.11.2019

Course Code: CSE 404 | **Time**: 9.30 AM to 10.30 AM

Course Name: SOCIAL NETWORK ANALYSIS | **Max Marks**: 40

Program & Sem: B.Tech (CSE) & VII Sem | **Weightage**: 20%

Instructions:

(i) *Read the question properly and answer accordingly.*

(i) *Question paper consists of 3 parts.*

(ii) *Scientific and Non-programmable calculators are permitted.*

Part A [Memory Recall Question]

Answer all the Questions. Each question carries six marks. **(3Qx4M=12M)**

1. Explain Cliques with relevant example. (C.O.NO.2)[Comprehension]
2. Explain k-clique and k-club reachability in a network graph. (C.O.NO.2)[Comprehension]
3. Explain the concept of Affiliation networks. (C.O.NO.2)[Comprehension]

Part B [Thought Provoking Question]

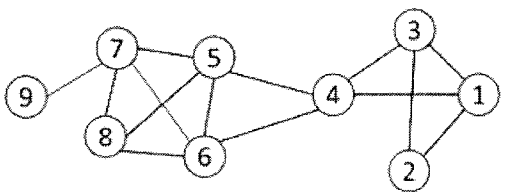
Answer both the Questions. Each question carries five marks. **(2Qx6M=12M)**

4. Explain the method of finding maximum clique in a given network with relevant example. (C.O.NO.2)[Comprehension]
5. Discuss the Edge Betweenness and vertex similarity, name 2 similarity measure. (C.O.NO.2)[Comprehension]

Part C [Problem Solving Questions]

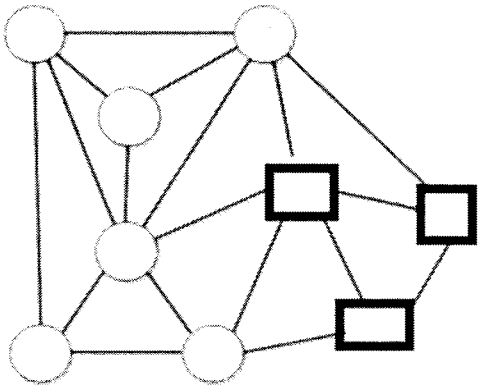
Answer both the Questions. Each question carries six marks. **(2Qx6M=12M)**

6. Apply the Clique Percolation Method (CPM) algorithm to detect at least two communities from within the following network. Assume the input parameter k = 3. (C.O.NO.2)[Application]



7. Write the homophily test algorithm. Apply the Homophily Test algorithm to check for any evidence of homophily in the following heterogeneous network graph.

(C.O.NO.2)[Application]



SCHOOL OF ENGINEERING

TEST – II

Semester: VII
Course Code: CSE 404
Course Name: Social Network Analytics
Program & Sem: B.Tech & VII

Date: 20-11-2019
Time: 9:30 to 10:30 AM
Max Marks: 40
Weightage: 20%

Extract of question distribution [outcome wise & level wise]

Q.NO	C.O.NO	Unit/Module Number/Unit /Module Title	Memory recall type [Marks allotted] Bloom's Levels			Thought provoking type [Marks allotted] Bloom's Levels			Problem Solving type [Marks allotted]			Total Marks
			K	C	A	K	C	A	K	C	A	
1	CO2	Module 2	-	04	-							4
2	CO2	Module 2	-	04	-							4
3	CO2	Module 3	-	04	-							4
4	CO2	Module 2				-	06	-				6
5	CO2	Module 3				-	06	-				6
6	CO2	Module 2							-	-	08	8
7	CO2	Module 3							-	-	08	8
	Total Marks			12			12				16	40

I hereby certify that all the questions are set as per the above guidelines. [MR .SATHISHI KUMAR]

Annexure- II: Format of Answer Scheme

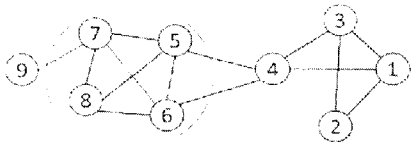
SCHOOL OF ENGINEERING

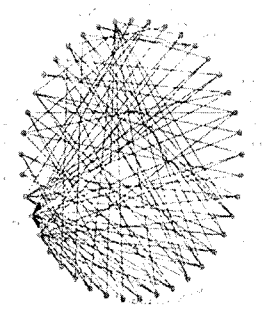
SOLUTION

Semester: VII
Course Code: CSE 297
Course Name: Social Network Analytics
Program & Sem: B.Tech & VII

Date: 20-11-2019
Time: 9:30 to 10:30 AM
Max Marks: 40
Weightage: 20%

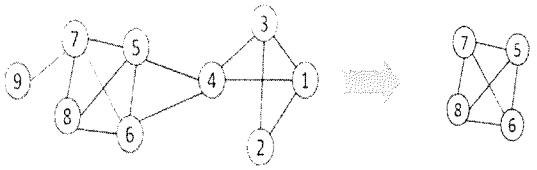
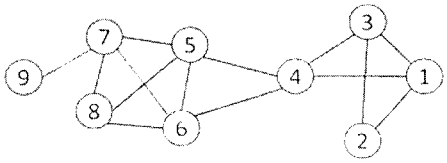
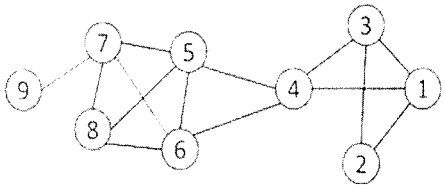
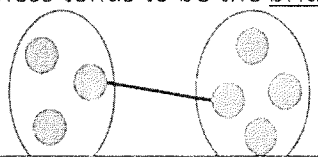
Part A (3Q x 4M = 12 Marks)

Q No	Solution	Scheme of Marking	Max. Time required for each Question
1	• Clique: a maximum complete subgraph in which all nodes are adjacent to each other  Nodes 5, 6, 7 and 8 form a clique • NP-hard to find the maximum clique in a network • Straightforward implementation to find cliques is very expensive in time complexity	**Definition=1** **Graph =2** **Solution=1**	6 min

2	**Reachability : k-clique, k-club** • Any node in a group should be reachable in k hops • k-clique: a maximal subgraph in which the largest geodesic distance between any two nodes <= k • k-club: a substructure of diameter <= k 2 4 1 6 3 5 Cliques: {1, 2, 3} 2-cliques: {1, 2, 3, 4, 5}, {2, 3, 4, 5, 6} 2-clubs: {1,2,3,4}, {1, 2, 3, 5}, {2, 3, 4, 5, 6} • Commonly used in traditional SNA • Often involves combinatorial optimization	**Describing =1** **Graph =1** **Solution=2**	6 min
3	**Affiliation Networks: a.k.a. two-mode network**  Affiliation networks are examples of a class of graphs called bipartite graphs Namely, nodes can be divided into two sets in such a way that every edge connects a node in one set to a node in the other set. (In other words, there are no edges joining a pair of nodes that belong to the same set; all edges go between the two sets.) Affiliation networks represents the participation of a set of people in a set of foci	**Definition=2** **Example=2**	6 min

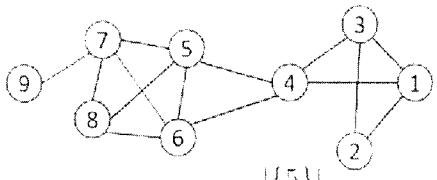
Part B (2Q x 6M = 12 Marks)

Q No	Solution	Scheme of Marking	Max. Time required for each Question

4	## Maximum Clique Example  • Suppose we sample a sub-network with nodes {1-9} and find a clique {1, 2, 3} of size 3 • In order to find a clique >3, remove all nodes with degree <=3-1=2 – Remove nodes 2 and 9 – Remove nodes 1 and 3 – Remove node 4	**Example=3** **Algorithm=3**	8 min
5	## Clustering based on Vertex Similarity • Apply k-means or similarity-based clustering to nodes • Vertex similarity is defined in terms of the similarity of their neighborhood • Structural equivalence: two nodes are structurally equivalent iff they are connecting to the same set of actors Nodes 1 and 3 are structurally equivalent; So are nodes 5 and 6.  • Structural equivalence is too restrict for practical use. ## Edge Betweenness • The strength of a tie can be measured by edge betweenness • Edge betweenness: the number of shortest paths that pass along with the edge  The edge betweenness of e(1, 2) is 4 (=6/2 + 1), as all the shortest paths from 2 to {4, 5, 6, 7, 8, 9} have to either pass e(1, 2) or e(2, 3), and e(1,2) is the shortest path between 1 and 2 • The edge with higher betweenness tends to be the bridge between two communities. 	**Vertex Similarity =2** **Edge Betweenness=2** **Similarity Measures=2**	10

Vertex Similarity

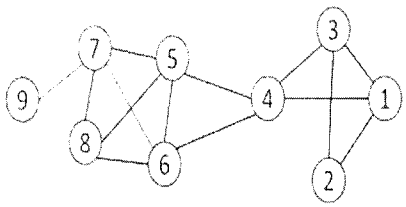
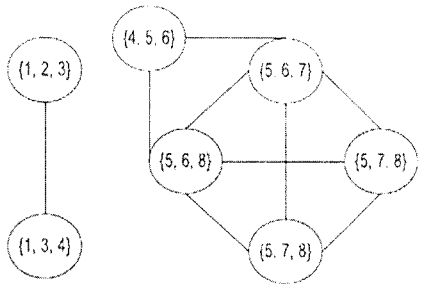
- Jaccard Similarity $Jaccard(v_i, v_j) = \frac{|N_i \cap N_j|}{|N_i \cup N_j|}$
- Cosine similarity $Cosine(v_i, v_j) = \frac{|N_i \cap N_j|}{\sqrt{|N_i| \cdot |N_j|}}$

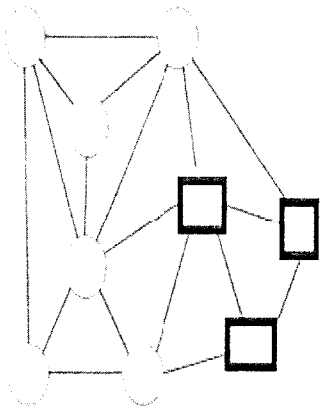


$$Jaccard(4,6) = \frac{|\{5\}|}{|\{1,3,4,5,6,7,8\}|} = \frac{1}{7}$$

$$cosine(4,6) = \frac{1}{\sqrt{4 \cdot 4}} = \frac{1}{4}$$

Part C (2Q x 8M = 16 Marks)

Q No	Solution	Scheme of Marking	Max. Time required for each Question
6	CPM Example  Cliques of size 3: {1, 2, 3}, {1, 3, 4}, {4, 5, 6}, {5, 6, 7}, {5, 6, 8}, {5, 7, 8}, {6, 7, 8} Communities: {1, 2, 3, 4} {4, 5, 6, 7, 8} 	**Algorithm =4** **Example=4**	12 min

7	**Homophily Test:** If the fraction of heterogeneous (cross-gender) edges is significantly less than $2pq$ then there is evidence for homophily  Cross-gender edges: 5 of 18 $p = 6/9 = 2/3$ $q = 3/9 = 1/3$ If no homophily, # of cross-gender edges should be $2pq = 4/9$ 8 out of 18 ∴ Evidence of homophily	**Algorithm=4** **Test=4**	12 min

SCHOOL OF ENGINEERING

END TERM FINAL EXAMINATION

Semester: Odd Semester: 2019 - 20
Course Code: CSE 404
Course Name: SOCIAL NETWORK ANALYSIS
Program & Sem: (CSE/ECE/EEE) & VII (OE-II)

Date: 30 December 2019
Time: 9:30 AM to 12:30 PM
Max Marks: 80
Weightage: 40%

Instructions:

(i) Read the question properly and answer accordingly
(i) Question paper consists of 3 parts.
(ii) *Scientific and Non-programmable calculators are permitted.*

Part A [Memory Recall Questions]

Answer all the Questions. Each Question carries 3 marks. **(5Qx3M=15M)**

1. Define 'degree' of a node in a network graph with proper example. (C.O.No.1) [Knowledge]
2. State the differences between walks, trails and paths in a network graph. (C.O.No.1) [Knowledge]
3. Describe the concept of Affiliation networks. (C.O.No.2) [Knowledge]
4. State the differences between Content based recommendation and Collaborative filtering with relevant examples. (C.O.No.3) [Knowledge]
5. List the challenges of recommender systems. . (C.O.No.3) [Knowledge]

Part B [Thought Provoking Questions]

Answer all the Questions. Each Question carries 5 marks. **(7Qx5M=35M)**

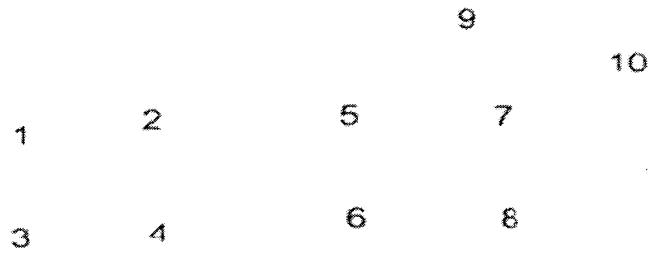
6. Explain the concept of Eigenvector centrality of a node in a network graph. Describe the steps to compute the same. (C.O.No.1) [Comprehension]
7. Discuss Cosine and Jaccard similarity measures with relevant example. (C.O.No.2) [Comprehension]
8. Explain the method of finding maximum clique in a given network with relevant example. (C.O.No.2) [Comprehension]
9. Discuss the Schelling model of segregation with proper illustrations. State its utility.

(C.O.No.3) [Comprehension]

Part C [Problem Solving Questions]

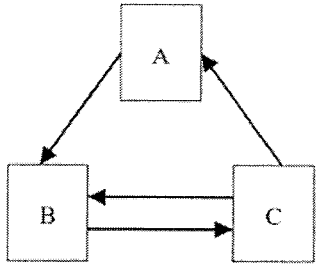
Answer all the Questions. Each Question carries 10 marks. **(3Qx10M=30M)**

13. Apply the Clique Percolation Method (CPM) algorithm to detect at least two communities from within the following network. Assume the input parameter k = 3.



(C.O.No.2) [Application]

14. Compute PageRank for pages A, B and C in the following network. Assume Initial rank of each page to be 1 and damping factor d = 0.85. (C.O.No.3) [Application]



15. Let us consider the following dataset as ratings of 3 Books B1, B2, B3 by 4 readers R1, R2, R3, R4. Using the collaborative filtering algorithm

Book No	Readers	Books	rating
241	R1	B1	3
222	R1	B3	2
276	R2	B1	5
273	R2	B2	3
200	R3	B1	2
229	R3	B2	3
231	R3	B3	1
239	R4	B2	3

END TERM FINAL EXAMINATION

Extract of question distribution [outcome wise & level wise]

Q.NO.	C.O.NO (% age of CO)	Unit/Module Number/Unit /Module Title	Memory recall type [Marks allotted] Bloom's Levels	Thought provoking type [Marks allotted] Bloom's Levels	Problem Solving type [Marks allotted]	Total Marks
			K	C	A	
1	CO1	Module 1	03			03
2	CO1	Module 1	03			03
3	CO2	Module 3	03			03
4	CO3	Module 4	03			03
5	CO3	Module 4	03			03
6	CO1	Module1		05		05
7	CO2	Module 2		05		05
8	CO2	Module 2		05		05
9	CO2	Module 3		05		05
10	CO2	Module 3		05		05
11	CO3	Module 4		05		05
12	CO3	Module 4		05		05
13	CO2	Module 2			10	10
14	CO3	Module 4			10	10
15	CO3	Module 4			10	10
	Total Marks		15	35	30	80

I hereby certify that all the questions are set as per the above guidelines.

Faculty Signature:

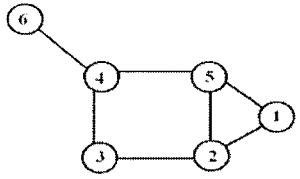
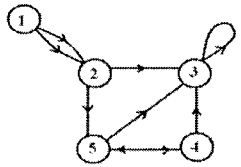
Reviewer Commend:

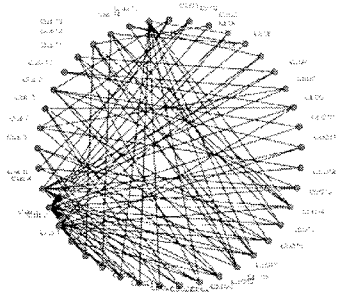
REACH GREATER HEIGHTS

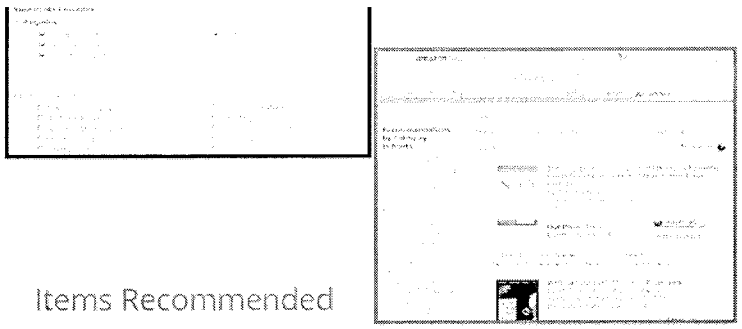
Semester: Odd Sem. 2019-20
Course Code: CSE 404
Course Name: SOCIAL NETWORK ANALYTICS
Program & Sem: B-TECH & VII

Date: 30.12.2019
Time: 3 HRS
Max Marks: 80
Weightage: 40%

Part A **(5Q x 3M = 15Marks)**

Q No	Solution	Scheme of Marking	Max. Time required for each Question
1	Degree (Un Directed Graphs) Number of edges incident on a node The degree of 5 is 3  Degree (Directed Graphs) In degree: Number of edges entering Out-degree: Number of edges leaving Degree = indeg + outdeg outdeg(1) = 2 indeg(1) = 0 outdeg(2) = 2 indeg(2) = 2 outdeg(3) = 3 indeg(3) = 4 	1.5 for undirected graph + 1.5 for undirected graph	5 Mins

	Walk (nodes and edges can be repeated): B -> A -> D -> A (what is passed through the network goes through the same node A twice, and through edge 3 twice) Trail (edges can not be repeated but nodes can): A -> D -> E -> A -> B (what is passed through the network does not use the same edge twice, but passes through node A twice) Path (no edges/nodes can be repeated): B -> A -> E -> D (no edge or node is repeated, what is passed through the network goes through the unique sequence of nodes and edges)		
3	Affiliation Networks: a.k.a. two-mode network  Affiliation networks are examples of a class of graphs called bipartite graphs. Namely, nodes can be divided into two sets in such a way that every edge connects a node in one set to a node in the other set. (In other words, there are no edges joining a pair of nodes that belong to the same set; all edges go between the two sets.) Affiliation networks represents the participation of a set of people in a set of foci.	**1.5 for definition + 1.5 for example**	**5 Mins**
4	**Assumption:** a user's interest should match the description of the items that the user should be recommended by the system. – The more similar the item's description to that of the user's interest, the more likely that the user finds the item's recommendation interesting. **Goal: find the similarity between the user and all of the existing items is the core of this type of recommender systems** **Content-based Recommendation Algorithm** 1. Describe the items to be recommended 2. Create a profile of the user that describes the types of items the user likes 3. Compare items with the user profile to determine what to recommend	**1.5 for Content Based Filtering + 1.5 for Collaborative Filtering**	**5 Mins**



Collaborative Filtering

- Match people with similar interests as a basis for recommendation.
  1) Many people must participate to make it likely that a person with similar interests will be found.
  2) There must be a simple way for people to express their interests.
  3) There must be an efficient algorithm to match people with similar interests.

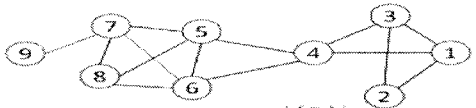
Example of CF MxN Matrix with M users and N items (An empty cell is an unrated item)

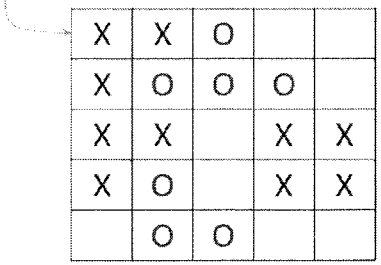
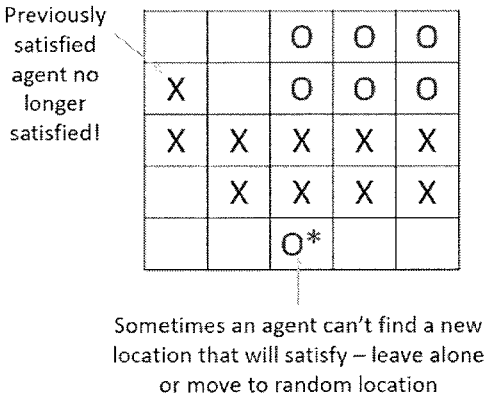
Items / Users	Data Mining	Search Engines	Data Bases	XML
Alex	1		5	4
George	2	3	4	
Mark	4	5		2
Peter			4	5

5	Challenges of Recommender System • **The Cold Start Problem** – Recommender systems use historical data or information provided by the user to recommend items, products, etc. – When user join sites, they still haven't bought any product, or they have no history. – It is hard to infer what they are going to like when they start on a site. • **Data Sparsity**	**3 marks for writing five challenges**	**5 Mins**

explanation on why these items are recommended

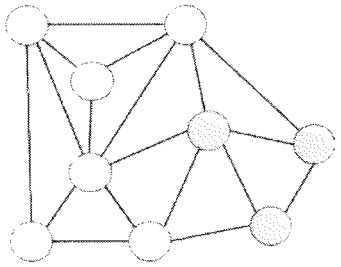
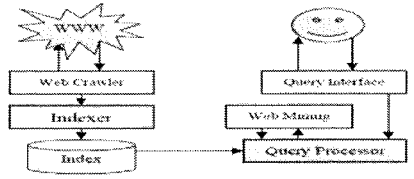
Part B **(7Q x 5M = 35 Marks)**

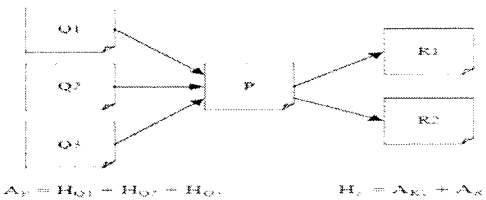
Q No	Solution	Scheme of Marking	Max. Time required for each Question
6	Problem: what if the node has high degree but all of its connections are not well-connected to the rest of the network. Shouldn't centrality be defined by how well-connected your connections are? Eigenvector centrality = A node is central to the extent that the node is connected to others who are central. Eigenvector centrality measure is based on walks. • Starts by assigning centrality score of 1 to all nodes (vi = 1 for all i) • Recomputes scores of each node as weighted sum of centralities of all nodes in a node's neighborhood: vi = xijvj • Normalizes v by dividing each value by the largest value • Repeat steps 2 and 3 until values of v stop changing.	**2.5 for definition + 2.5 for algorithm**	**13 Mins**
7	**Vertex Similarity** • Jaccard Similarity $Jaccard(v_i, v_j) = \frac{|N_i \cap N_j|}{|N_i \cup N_j|}$ • Cosine similarity $Cosine(v_i, v_j) = \frac{|N_i \cap N_j|}{\sqrt{|N_i| \cdot |N_j|}}$  $Jaccard(1,6) = \frac{|\{5\}|}{|\{1,3,4,5,6,7,8\}|} = \frac{1}{7}$ $cosine(1,6) = \frac{1}{\sqrt{4 \cdot 4}} = \frac{1}{4}$	**2.5 for equation + 2.5 for example**	**13 Mins**
8	**Finding the Maximum Clique** In a clique of size k, each node maintains degree >= k-1 – Nodes with degree < k-1 will not be included in the maximum clique Recursively apply the following pruning procedure – Sample a sub-network from the given network, and find a clique in the sub-network. – Suppose the clique above is size k, in order to find out a *larger* clique, all nodes with degree <= k-1 should be removed. Repeat until the network is small enough Many nodes will be pruned as social media networks follow a power law distribution for node degrees	**2.5 for definition + 2.5 for example**	**15 Mins**

	• Suppose we sample a sub-network with nodes {1-9} and find a clique {1, 2, 3} of size 3 • In order to find a clique >3, remove all nodes with degree <=3-1=2 – Remove nodes 2 and 9 – Remove nodes 1 and 3 – Remove node 4		
9	There is a population of individuals of two types. Each individual wants to have at least *t* other agents of its own type as neighbors. Unsatisfied individuals move in a sequence of rounds as follows in each round, in a given order, each unsatisfied moves to an unoccupied cell where it will be satisfied (details can differ with similar qualitative behaviour). These new locations may cause different individuals to be unsatisfied, leading to a new round of movement. Agents placed randomly in grid  Agents are **satisfied** with their location in the grid if at least *t* of their neighbors are agents of the same type. *t* = threshold Note that the grid is now more segregated than before  	**2 for definition + 2 for example + 1 for utility**	**15 Mins**

X	X	O			
X	O	O	O		
X	X		X	X	
X	O		X	X	
	O	O			

		O	O	O
X		O	O	O
X	X	X	X	X
	X	X	X	X
		O*		

10	*people of the same sort or with the same tastes and interests will be found together.* • Your friends are more similar to you in age, race, interests, opinions, etc. than a random collection of individuals • **Homophily**: principle that we tend to be similar to our friends **Homophily Test**: If the fraction of heterogeneous (cross-gender) edges is significantly less than $2pq$ then there is evidence for homophily  Cross-gender edges: 5 of 18 $p = 6/9 = 2/3$ $q = 3/9 = 1/3$ If no homophily, # of cross-gender edges should be $2pq = 4/9 =$ 8 out of 18 ∴ Evidence of homophily	**2.5 for definition+ 2.5 for example**	**13 Mins**
11	With the rapid growth of WWW most of the users use information retrieval tools like search engines to find information from the web.  There are tens and hundreds of search engines available but some are popular like Google, Yahoo, Bing etc., because of their crawling and ranking methodologies. The search engines download, index and store hundreds of millions of web pages. They answer tens of millions of queries every day. So Web mining and ranking mechanism becomes very important for effective information retrieval. Before presenting the pages to the user, a ranking mechanism is done by the search engines to present the most relevant pages at the top and less relevant ones at the bottom.	**3 for architecture + 2 for illustration**	**12 Mins**
12	HITS • Hyperlink Induced Topic Search(HITS) Algorithm [13] ranks the web page by processing in links and out links of the web pages. In this algorithm a web page is named as authority if the web page is pointed by many hyperlinks and a web page is names as HUB if the page points to various hyperlinks. Authorities and hubs are illustrated in Figure • Hubs and authorities are assigned respective scores. Scores are computed in a mutually reinforcing way; an authority pointed to by several highly scored hubs should be a strong authority while a hub that points to several highly scored authorities should be a popular hub. • Let ap and hp represent the authority and hub scores of page p, respectively. B (p) and I (p) denote the set of referrer and reference pages of page p, respectively. The scores of hubs and authorities are calculated as follows; $$ap = \sum_{q \in B(p)} hp$$	**2.5 for hubs + 2.5 for authorities**	**14 Mins**

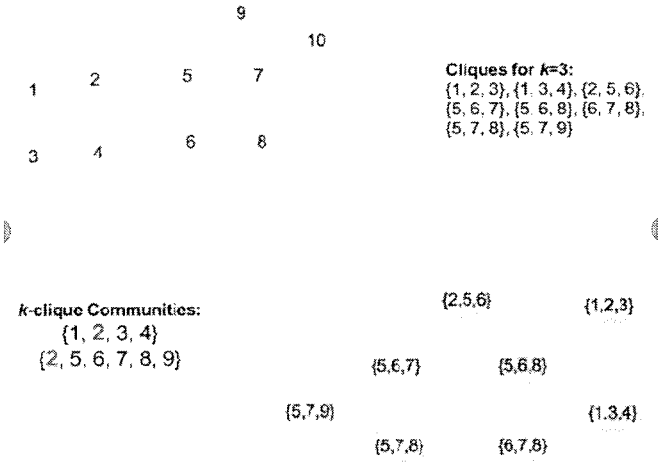


$A_P = H_{Q1} + H_{Q2} + H_{Q3}$ $H_P = A_{R1} + A_{R2}$

HITS is a purely **link-based** algorithm. It is used to rank pages that are retrieved from the web, based on their textual contents to a given query. Once these pages have been assembled, the HITS algorithm ignores textual content and focuses itself on the structure of the web only.

Part C

(3Q x 10M = 30Marks)

Q No	Solution	Scheme of Marking	Max. Time required for each Question
13	• CPM is such a method to find overlapping communities – **Input** • A parameter k, and a network – **Procedure** • Find out all cliques of size k in a given network • Construct a clique graph. Two cliques are adjacent if they share k-1 nodes • Each connected components in the clique graph form a community 	**5 for Algorithm + 5 for Calculation**	**20 Mins**

	So PR(A) = 0.15 + 0.85 × 1/2 PR(B) = 0.15 + 0.85 × PR(A) + 1/2 PR(C) = 0.15 + 0.85 × PR(B) By solving the above system of linear equations, we get PR(A) = 0.575 PR(B) = 1.13 PR(C) = 1.11	**6 for Calculation**	
15	i) B1 B2 B3 R1 3 ? 2 R2 5 3 ? R3 2 3 1 R4 ? 3 2 ii) B1 B2 B3 B1 1 1 0.99 B2 1 1 0.78 B3 0.99 0.78 1 iii) B1 B2 B3 R1 3 **2.56** 2 R2 5 3 **4.11** R3 2 3 1 R4 **2.5** 3 2	**3 for Matrix + 3 for Book-Book Similarity Matrix + 4 for predicting ratings**	**20 Mins**