2

SL.NO:CAA 212772

SRI SIVASUBRAMANIYA NADAR COLLEGE OF ENGINEERING

(An Autonomous Institution, Affiliated to Anna University, Chennai)
Rajiv Gandhi Salai (OMR), Kalavakkam – 603 110

THEORY EXAMINATIONS

Register Number	205001085		
Name of the Student	Sabarivasan. V		
Degree and Branch	B.E. CSE	Semester	VII
Subject Code and Name	UCS1722 Social Network Analysis		
Assessment Test No.	II	Date	17/10/2023

Details of Marks Obtained

Part A		Part B				Part C			
Question No.	Marks	Question No.	(a) Marks	(b) Marks	Total Marks	Question No.	(a) Marks	(b) Marks	Total Marks
1	2	7	5			10	9		
2	1								
3	1	8	6			11			
4	2					12	7		
5	2	9	6			13			
6	2								
Total (A)		Total (B)				Total (C)			

Grand Total (A+B+C)				Marks (in words)			
Signature of Faculty							

PART-C

(10) Optimization based algorithms for detecting communities in networks.

⇒ They are classified into 2 types.

(i) Spectral Algorithms

(ii) Local Search Algorithms.

(i) Spectral Algorithms

⇒ In these algorithms, the goodness of the partition on networks are done using cuts.

⇒ The basic spectral algorithm will be spectral bi-partition.

⇒ The cuts are made at appropriate edges to increase the goodness of partition and lead to detection of communities.

⇒ Sometimes this algorithm may not give the optimised algorithm for large networks. Hence different cuts such as average, ratio or normalised cuts can be made.

⇒ They basically provide the local maximum solution on an approximated time.

⇒ They are classified as,

(i) Kernighan-Lin Algorithm

(ii) Fast-Newman Algorithm

(iii) Girvan-Newman Algorithm

(iv) Potts Algorithm

(i) Kernighan-Lin Algorithm

⇒ It provides solution based on an evaluation function which is defined as the difference between the number of intra-nodes and inter nodes connections.

⇒ The nodes are either swapped or moved by decreasing the evaluation function.

⇒ The algorithm terminates if no change in evaluation function.

⇒ Time complexity = $O(n^2)$

⇒ Prior knowledge about the nodes and connections are need to be known already.

(ii) Fast-Newman Algorithm

⇒ Time complexity = $O(mn)$.

⇒ At the algorithm beginning, initially single node is present, and slowly best merges are made among the nodes in such a way, at last a single community will be left off.

⇒ This type of best-merge tactics leads to dendrogram construction based on the modularity function Q.

⇒ $Q = \sum e_{ij} - a_i^2$ where e_{ij} is the number of paths between node 'i' and node 'j'.

$a_i = \sum e_{ij}$

Clauset Arnoud Algorithm

⇒ The algorithm is working based on maximum modularity function value.

⇒ The commonly used technique here is Annealing process which is basically a global stochastic optimization algorithm.

⇒ This is continued by applying randomness over the network to start annealing process.

⇒ The node chosen is given by the 'metropolis' modularity value,

$$p = \begin{cases} 1, & C_{t+1} \leq C_t \\ e^{-(C_{t+1} - C_t)/T}, & C_{t+1} > C_t \end{cases}$$

where $C_t = -Q_t$ at a given time $t+1$ and temperature Q_t.

(iv) Potts Algorithm:

⇒ Each node have different spins with different value.

⇒ Nodes which have stable spin are chosen.

⇒ The stable chosen community will have same spin values which denotes minimum energy is used.

⇒ A community having same spin values for all nodes is said to be a single community entirely.

(12)

Simplify network for Analysis

SSN

(i) Filter using degrees

The nodes can be filtered using degrees to find the most of the connections in a network.

(ii) Applying centralities:

Apply different centrality based algorithms such as betweenness centrality, edge centrality to distinguish the nodes in a network.

(iii) Reducing dimensions:

The dimensions of the network can be reduced to a simpler form using different techniques and algorithms without changing the topology of the network.

(iv) Detect communities

SSN

Detecting communities from a graph or network could be easier to differentiate the common traits and properties presented in each community.

Networkx Code

```
import networkx as nx

connections = [('A','B'), ('B','C'), ('C','D')]

for i in connections:
    G.add_edge(i)  # add edges

# apply degree centrality on network.
for node in Graph.nodes():
    values.append(Graph.degree(node))

# return the degree centrality result.
return values.
```

Detect most influential node in network

(i) Apply degree centrality:

Nodes which have high degrees are may be considered as influential nodes since most of the connections are linked to those nodes.

(ii) Node connecting communities

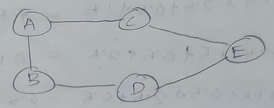
Detect the node which connects most of the communities together since they act as bridges.

(iii) Nodes with higher values

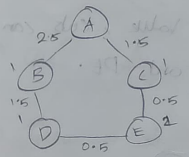
Nodes with higher values may also be detected as influencial node as it might hold the control of the major side of the network.

PART-B

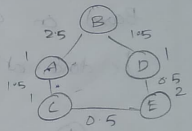
(9) Girvan-Newmans Algorithm



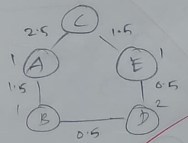
BFS at A



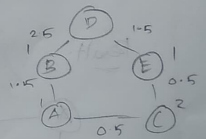
BFS at B



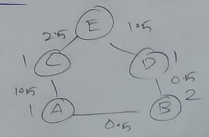
BFS at C



BFS at D

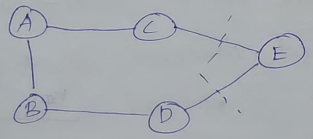


BFS at E

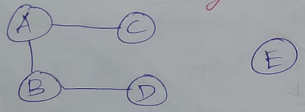


- AB = 1.5 + 1.5 + 1.5 + 1.5 + 0.5 = 7.5
- AC = 1.5 + 1.5 + 2.5 + 0.5 + 1.5 = 7.5
- CE = 0.5 + 0.5 + 1.5 + 0.5 + 0.5 = 10.5
- BD = 1.5 + 1.5 + 0.5 + 2.5 + 0.5 = 6.5
- DE = 6.5 + 0.5 + 0.5 + 1.5 + 1.5 = 10.5

Based on the above value, cuts can be done at edge CE and DE.



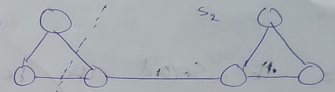
Result:



E

(8) Newman - Girvan modularity

Case 1:

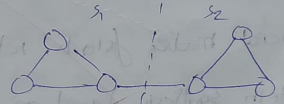


$$Q = \sum \left(\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right)$$

$$\Rightarrow Q_{S_1} = \frac{1}{7} - \left(\frac{4}{14} \right)^2 = 0.06$$

$$\Rightarrow Q_{S_2} = \frac{4}{7} - \left(\frac{10}{14} \right)^2 = 0.06$$

Case 2:



$$\Rightarrow Q_{S_1} = \frac{3}{7} - \left(\frac{7}{14} \right)^2 = 0.179$$

$$\Rightarrow Q_{S_2} = \frac{3}{7} - \left(\frac{7}{14} \right)^2 = 0.179$$

⇒ From the above 2 cases, the 2nd case has the high modularity value.

⇒ Hence the second case network can be considered for partition.

PART - A

(1) Non-directed relations in social network

→ Family member relations.

→ Children playing games with other children

→ Social media friends network

→ Protein synthesis based on different organisms form a network which is based on different chemical reactions.

(2) ⇒ While children playing together, different children have different mindset to play different games and to play with different peoples

⇒ Here the relation could be children to games and children to his friends

(3) Detecting Communities

⇒ Detecting communities in a social network is important because different traits and properties are mixed in a network

⇒ Hence dividing them into communities with similar properties and traits will be useful for research purpose and other functionality.

(4) Modularity of Newman-Girvan Algorithm

$$Q = \sum_{i}^{n_s} \left(\frac{l_s}{m} - \left(\frac{d_s}{2m} \right)^2 \right)$$

n_s → number of edges in network

l_s → number of edges in a community 's'

d_s → degree sum of a community 's'

m → number of total edges.

left term → fractions of edges in community

right term → expected fractions of community

(5)

Optimization Based algorithms	Heuristics Based Algorithm
They find solutions by optimizing the function and find the optimized result.	Search for local or global maximum in a network which gives approximate optimized solution.
Difficult in applying on larger networks.	Can be applied for larger network
Eg: Spectral Algorithm.	Eg. Wu-Huberman Algorithm

(6) Networkx code

```
import networkx as nx
for node in Graph.nodes():
    if degree(node) > 2:
        remove_list.append(node)
return [nodes for nodes in Graph.nodes() if nodes not in remove_list]
```

(7)

(i) Local Definition

(ii) Global Definition

(iii) Vertex Similarity

(i) Local Definition. It is defined as nodes and edges are connected with betweenness, centrality.

Clique: It is a maximal subgraph of a network which nodes are adjacent to other.

(ii) Global definition

It is defined as nodes connected with other communities using different aspects and properties.

(iii) Vertex Similarity

→ It is defined as the nodes and edges are grouped based on the similar properties of vertices on a graph.

→ This leads to dendrogram construction by drawing horizontal cuts over the larger network.